MOTION-DEBLURRING MECHANISMS OF HUMAN VISUAL PERCEPTION

Ari Kullervo Pääkkönen

Doctor of Philosophy
University of Edinburgh
1993

Abstract

The temporal integration period of the visual system for both stationary and moving objects is known to be over 100 ms in daylight. One might expect from this that motion blur should considerably degrade our percepts of moving objects. However, we are able to see moving objects with clarity. Some authors have suggested that there are special motion-deblurring mechanisms to prevent motion degradation. The proposed mechanisms are motion-tuned: integration follows the motion path of the object. The aim of this work was to study the properties of visual motion blur and their implications for possible deblurring mechanisms.

In the main experiment, blur discrimination thresholds for moving, Gaussian-blurred edges were measured in order to separate the effects of motion blur and static spatial blur. The data were modelled by assuming a linear transform of the physical image of the edge to its neural representation. The results show that motion produces equivalent spatial blur. The velocity dependence of this blur is linear, and its extent can be predicted by a temporal impulse response with a standard deviation of about 5 ms in normal room light, which gives an estimate of about 20 to 25 ms for the blur-producing motion integration time. Supported by the results of two-dot resolution experiments, it is proposed that this blur results from camera-like summation and not from the use of larger spatial filters for moving than for stationary objects.

When the author repeated the main experiment six months later after many other experiments mainly with small reference blurs, changes in thresholds and model parameters were found that indicate low-level learning and neural plasticity in the blur discrimination system. An even more surprising result came from the experiment where the contrast polarity in relation to the direction of motion was changed opposite to that in the main experiment. Learning effects for moving edges were not transferred from one polarity to the other, indicating that spatial analysis of moving objects may be served by two separate subsystems, possibly related to the on and off systems of the visual pathway.

The estimated amount of motion blur corresponds to only about a fifth of what would be expected from the summation time of about 100 ms. To explain the difference, some kind of blur-free integration

mechanism is needed. It is argued here supported by theoretical considerations and evidence from a spatial interval discrimination experiment that this mechanism does not have to be motion-tuned, and an alternative model for temporal integration of moving objects is presented. In this model, integration happens in two phases. The first phase is a camera-like exposure phase that always produces motion blur. The images produced by the first phase are then integrated by the second phase which is translation-invariant: it uses some mechanism that can take into account changes of object location even if they are random. This phase does not produce motion blur, but it does not remove it either. Instead, it increases the signal-to-noise ratio or the statistical significance in discriminating the features in the image. The mechanism could be based on cross-correlation of successive images. A related model that uses correlation principles to form a simultaneous representation of the form and the velocity of a moving object is also proposed.

Preface

No part of this thesis, in any form, has been submitted for any other degree or professional qualification. To the best of my knowledge, none of the experiments nor the modelling reported here has been previously carried out by any other investigator. All experimental work was carried out in the Laboratory for Neuroscience at the Department of Pharmacology, University of Edinburgh. Parts of the thesis have been accepted for publication in the Journal of the Optical Society of America A (Pääkkönen and Morgan, "The effects of motion upon blur discrimination").

Acknowledgements

This work was carried out in the Laboratory for Neuroscience at the Department of Pharmacology, University of Edinburgh, during the years 1990-1993. I wish to thank Professor John Kelly, Head of the Department of Pharmacology, for placing the facilities of the department at my disposal.

I want to express my sincere gratitude to my supervisor, Professor Michael Morgan, for his great help in making it possible for me to study in Edinburgh, and for his encouragement and advice in all phases of my work.

I am grateful to my other supervisor, Dr. David Willshaw, for many valuable discussions and suggestions during the study.

My thanks go to my fellow workers at the Laboratory for Neuroscience, Dr. Linda Bowns, Dr. James Tresilian, and Mr. Peter Lunn, B.Sc., for fruitful discussions. Special thanks go to Dr. David Keeble for discussions, and for reading the manuscript and giving critical comments. I also thank Ms. Monika Stedul, B.Sc., and Mr. Richard Osborne for serving as observers in the experiments.

I wish to thank my former teachers at the Department of Applied Physics, University of Kuopio, and in the University Hospital of Kuopio for many skills that helped me in this study, and especially for teaching me the importance of modelling.

My warm thanks go to the whole personnel of the Department of Clinical Neurophysiology, Kuopio University Hospital, for their friendly support.

This work was financially supported by the Academy of Finland, the Faculty of Medicine of the University of Edinburgh, the Department of Clinical Neurophysiology of Kuopio University Hospital and the Overseas Research Students Awards Scheme. The support is highly appreciated.

Finally, my loving thanks go to my dear wife Kirsti and my sons Jyri and Jere for their love, patience and support during this study.

Ari Pääkkönen
June, 1993

Contents

1. Introduction

The perception of motion is arguably one of the most fundamental abilities of the visual system, and an important part of it is the shape analysis of moving objects. One would expect from the poor temporal resolution of the visual system - the temporal integration period for both stationary (Barlow, 1958) and moving objects (Burr, 1981; Burr & Ross, 1986) has been estimated to be over 100 ms in daylight - that motion blur should degrade this shape analysis. It is rather surprising that some of the finest pattern discriminations or hyperacuities, like for instance vernier acuity (Westheimer & McKee, 1975) and stereoscopic acuity (Westheimer & McKee, 1978), are little affected by motion up to velocities of several degrees of arc per second. The effect of motion blur on our subjective impression is similar: we are not normally conscious of motion blur on moving objects (Burr, 1980). These facts have stimulated several explanations.

There are two main approaches to explain the absence of motion degradation in pattern discrimination tasks. Some authors (Burr, Ross & Morrone, 1986; Anderson & van Essen, 1987) have suggested that the visual system uses special motion-deblurring or blur prevention mechanisms in the analysis of the shape of moving objects. The basic principle of these mechanisms is to restore positional acuity by using the information about the temporal delay at which different photoreceptors have been stimulated. The other approach (Morgan & Benton, 1989; Morgan, 1991) is against any general motion-deblurring mechanism. It assumes that image motion does introduce spatial blur but claims that motion blur remains undetected because its amount is smaller than that of the spatial blur already present in the internal representation of the object.

Burr, Ross and Morrone (1986) used a masking technique to measure the joint spatiotemporal tuning functions of motion detectors, and assuming linearity at threshold, they calculated the spatiotemporal receptive fields of the detectors by inverse Fourier transform of the data. The fields obtained comprise alternating ridges of opposite polarity, elongated in space-time along the preferred velocity axis of the detector. They used the concept of a spatiotemporal receptive field to explain many phenomena of motion perception, one of which is the blur-free perception of moving

objects. The essence of the explanation is simple: a spatiotemporally oriented receptive field integrates a moving object not statically, but along its path of motion. If the orientation of the receptive field corresponds exactly to the velocity of the object, integration produces no smear.

Anderson and van Essen (1987) proposed a general strategy for dynamic control in the mapping of retinal output onto higher levels. They argued that dynamic switching principles would be desirable in many computational visual problems. One of the examples was the prevention of blurring of moving images. Dynamic linear shifts are used to align input and output maps without loss of local spatial relationships. The shifts are produced in increments at consecutive relay stages and controlled at each stage by lateral inhibitory connections. A descending feedback loop signalling the retinal velocity of the object is needed to give a measure of the alignment to the shift control circuitry. Both the shifter circuit and the spatiotemporal receptive field hypothesis have the same basic principle: integration follows the motion path of the object. In this respect, both models are basically motion-tuned.

Morgan and Benton (1989) found different motion sensitivities for vernier acuity and spatial interval acuity and regarded this as evidence against general motion-deblurring mechanisms. Their explanation was that spatial interval acuity is more sensitive to motion blur than vernier acuity because with closely placed parallel bars, motion blur causes the responses of the two bars to overlap, and their peaks become difficult to discriminate. On the other hand, however blurred the vertical vernier lines are by horizontal motion, they do not overlap on the retina. The explanation is sound, but nevertheless having different motion sensitivities for different hyperacuity tasks is not a sufficient argument against a general motion-deblurring mechanism. If two hyperacuity tasks have different sensitivities to static spatial blur, they will also have different sensitivities to motion. To study the issue properly, psychophysical tasks where the effects of static spatial blur and motion blur can be compared on a quantitative basis are needed.

This work was designed to study both the proposed motion-deblurring mechanisms and other plausible alternatives by investigating the properties of visual motion blur, especially the characteristics of temporal summation of moving objects.

2. Background

2.1. The visual pathway

Vision is a process for making inferences about the external physical world, based upon information contained in the spatial and temporal distribution of light coming into the eye. It solves its tasks so efficiently and seemingly effortlessly that it is easy to underestimate the complexity of the visual system. The primate visual cortex contains dozens of distinct areas in the cerebral cortex and several major subcortical structures (van Essen, Anderson & Felleman, 1992). The information processing strategies of the visual system include linear and nonlinear filtering, and coordinated use of different types of information.

2.1.1. Overview

The processing of the visual information starts when light enters the eye. The eye optics form a two-dimensional image of the three-dimensional world on the retina. The retina has more that 100 million light-sensitive photoreceptors: rods and cones. Rods are responsible for visual function at very low levels of light adaptation (scotopic vision). Cones enable us to see colours and fine detail at high levels of light adaptation (photopic vision). There are three types of cones with different spectral sensitivities: short-, medium- and long-wavelength of visible light. Combined, the cones provide our vision with a palette that can mix any colour.

The cones are most densely packed in the centre of the two degrees of visual arc wide foveal region, where they subtend about 30 seconds of arc. This value corresponds to approximately half the period of the highest possible spatial frequency in the retinal image predicted from the optical measures of the eye (Campbell & Gubisch, 1966). Thus, in the foveal centre or foveola the spacing of cones fulfils the requirement of the sampling theorem. However, cone spacing increases monotonically with eccentricity, and even 0.5° from the foveola the spacing is clearly coarser. On the other hand, the number of transmission lines from the retina (equivalent to the number of ganglion cells) is only about one hundredth of the

transmission, partly to avoid aliasing, partly to improve sensitivity and partly perhaps to diminish the input load of the central processing units (Barlow, 1981).

Photoreceptors transform light energy to electrochemical signals that are carried to bipolar cells by neurotransmitters. There are two major classes of bipolar cells: those that respond to increase of light (ON bipolars) and those that respond to decrease of light (OFF bipolars). ON- and OFF-centre ganglion cells receive their inputs from the corresponding kinds of bipolar cell. The functional role of the ON/OFF distinction is to widen the dynamical range of signal transmission (Schiller & Logothetis, 1990), i.e. to be a partial solution to the mismatch between the about ten decades wide range of visible luminances and less than fifty distinguishable signal levels of nerve fibres (Barlow, 1981). Horizontal and amacrine cells relay impulses laterally, processing visual information across the retina.

Gouras and Zrenner (1981) concluded that there are at least 11 types of ganglion cells with different centre-surround organization subserving in parallel any one small area of primate retina. The two major classes, ON- and OFF-centre cells, were originally described by Kuffler (1953) in the cat's retina. Hubel and Wiesel (1960) were the first to find substantially the same receptive field organization in primates. When the centre of their receptive fields is stimulated by light, the ON-centre cells increase their firing rate from the resting state. The centre is enclosed by an annular surrounding region of lower sensitivity, and the response is reduced by simultaneous stimulation of this surround. In OFF-centre cells the polarity of sensitivities is reversed. The shape of the function that describes how the position of an illuminated spot in the receptive field affects the output of the ganglion cell resembles a difference of two two-dimensional Gaussians with different standard deviations. Thus, this type of receptive fields are often described by difference of Gaussians or DOG functions. ON-centre ganglion cells are sometimes called "on centre-off surround cells" and one might conclude from this that the centre-surround organization of ganglion cell receptive fields is produced by the ON and OFF pathways of the retina. However, according to the present consensus, this is not the case: centre-surround antagonism is produced by the lateral connectivity of the amacrine and horizontal cell systems, and ON and OFF pathways have their own centre-surround organizations (Regan, 1989).

In both primates and cats a second grouping of ganglion cells involves parallel systems of tonic and phasic ganglion cells, each with its separate ON- and OFF-centre systems. Tonic cells give a prolonged or sustained response to prolonged stimulation, whereas phasic cells give only a transient response, acting almost like a differentiator or a high-pass filter in electronics. All phasic ganglion cells seem to lack colour coding. They are more common in the retinal periphery than in the fovea.

Nerve fibres from the ganglion cells form the optic nerve, through which impulses leave the eye. The fibres from each eye meet beyond the eye at the optic chiasm. The chiasm splits each optic nerve in half, forming two tracts. Fibres from the nasal halves of each retina cross over and project to the opposite hemisphere of the brain. Thus, messages from the same half of the visual field of both eyes reach the same hemisphere. Each hemisphere receives signals from only one half of the visual world, but the brain combines them into an integral perception. Bringing the information from the same part of the visual field from both eyes together is one prerequisite for stereoscopic vision.

After the optic chiasm the major part of the optic fibres in each tract go to the lateral geniculate nucleus (LGN), a relay station in each hemisphere. The LGN does not receive input only from the retina, but also from descending fibres conducting signals from the cortex, and these signals can modify the early processing of visual signals. The ON- and OFF-pathways remain distinct through the LGN and possibly even further. The receptive fields of the LGN cells are similar to those of the ganglion cells: they are monocular with circularly symmetrical centre-surround organization.

Most visual signals travel directly from the lateral geniculate nuclei to the primary visual cortex, also called the striate cortex because of its striped appearance. From there, nerve fibres project to a mosaic of extrastriate cortical visual areas. Thirty-two distinct cortical areas associated with visual processing have been described on the basis of anatomical, physiological and behavioural information (van Essen, Anderson and Felleman, 1992). Twenty-five of these are primarily visual in function. To date, 305 pathways interconnecting these areas have been identified. These figures correspond to the visual areas of the macaque monkey, a primate whose visual system is in many ways similar to that of humans.

The cells of the first cortical terminal are still monocular with concentric receptive fields. Beyond this stage many cells receive

input from both eyes and have elongated receptive fields. They respond well to a moving bar of some preferred orientation, but do not respond to a bar orthogonal to the preferred orientation, i.e. they are tuned to orientation. Hubel and Wiesel (1959; 1962; 1968) were the pioneers in studying the receptive field properties of the neurones in the visual cortex. Using flashing and drifting spots and thin lines as stimuli, they defined three classes of cells in the striate cortex of cat and monkey: simple, complex and hypercomplex cells. Today this classification and the receptive field properties described by them are known to be oversimplified, although some textbooks present them as the whole truth. Single neurones have also started to loose their almost thirty years unchallenged position as corner stones in explaining visual properties to more circuit-oriented concepts (Douglas & Martin, 1991). However, bearing this in mind, the descriptions given by Hubel and Wiesel are still giving us insight into the properties of the visual cortex.

According to Hubel and Wiesel, simple cells have receptive fields with separate excitatory and inhibitory regions. However, instead of being concentric, these antagonistic regions form parallel strips. They respond well to a moving bar when the bar is oriented parallel to the strips of the receptive field. Complex cells have many small interdigitated excitatory and inhibitory regions in their receptive fields. They respond only weakly to a stationary stimulus. Many complex cells are both direction specific and orientation selective: they respond best when a correctly oriented bar moves in the preferred direction. Thus, the cortical receptive fields are oriented in both spatial (x-y) and space-time (x-y-t) coordinates. Spatial orientation gives rise to orientation selectivity and spatiotemporal orientation to velocity tuning, i.e. direction and speed tuning. Cells in monkey LGN do not show direction selectivity or orientation tuning, so these properties of cortical cells must be created by cortical connectivity (Regan, 1989). Hypercomplex cells or end-stopped cells differ from the other classes in that the response starts to reduce if the stimulus bar becomes too long.

2.1.2. Parallel processing streams

Physical, anatomical and psychophysical studies have identified several parallel pathways or channels of information processing in the primate visual system. The ON- and OFF-pathways introduced in the previous chapter are one example of this parallelism. The ON- and OFF-pathways serve the same visual subprocesses, both carrying information essentially about the same variable. However, there are other distinctions based on pathways serving different subprocesses of vision, for instance form, colour, motion and depth. According to some investigators, different components of visual information processing are segregated into largely independent parallel channels (Livingstone & Hubel, 1987); others think that there is significant overlap across a number of paths and cortical areas (DeYoe & Van Essen, 1988; Schiller & Logothetis, 1990; van Essen, Anderson and Felleman, 1992). The two major processing streams are the parvocellular and the magnocellular pathways, the second of which is often been linked to motion analysis.

The two major processing streams originate in the retina. More than 80% of retinal ganglion cells are colour-opponent cells projecting to the parvocellular (P) layers of the lateral geniculate nucleus, whereas about 10% are broad-band cells projecting to magnocellular (M) layers of the LGN. Some investigators call parvo- and magnocellular channels colour-opponent and broadband channels, respectively, on the basis of their retinal ganglion cell properties. The properties of the cells of each stream, however, are very similar on both the retinal and LGN levels: The P cells have small receptive fields and their centre typically receives input from only one cone type, but the input to the surround is undifferentiated (Schiller & Logothetis, 1990). They respond in a sustained fashion to visual stimulation of low and moderate temporal frequencies (1 to 20 Hz) and have axons with medium conduction velocities. At any given eccentricity, the receptive fields of M cells are two to three times larger than those of P cells, thus providing only one third of the spatial resolution of P cells. They lack colour-opponency, with undifferentiated input from all cone types to both the centre and the surround of their receptive fields. They respond in a transient fashion to visual stimulation and have axons with high conduction velocities (Shapley & Perry, 1986). They also have shorter latencies than parvocellular cells, and they can

follow higher frequencies of stimulation, being optimized for temporal frequencies mainly from 5 to 20 Hz (Derrington & Lennie, 1984). M cells are more sensitive to luminance contrast than P cells, they respond well to stimuli below 10% contrast, with a contrast gain of about ten times higher than that of P cells (Kaplan & Shapley, 1986).

Cells in the magnocellular layers of the LGN project to layer 4B in V1 or visual area 1. From there most of the axons project to the midtemporal cortical area MT either directly or through the thick stripes (in cytochrome oxidase staining) of V2. Some axons project to V3. The outputs of the M stream are distributed from MT and V3 mainly to visual areas in the posterior parietal cortex. The P stream is divided into two within cortical areas V1 and V2. The P-I (parvo-interblob) stream receives most of the parvocellular LGN inputs and projects through the interblobs of V1 to the pale stripes or interstripes of V2. The P-B (parvo-blob) stream receives indirect input from the parvocellular LGN and direct inputs from the interlaminar layers of the LGN. It projects through the blobs of V1 to the thin stripes of V2. From V2 both the P-I and P-B streams project heavily to area V4, possibly remaining distinct within V4. The outputs of the P streams are distributed to subdivisions of inferotemporal cortex (Livingstone & Hubel, 1987; DeYoe & Van Essen, 1988).

The receptive field properties of cortical cells bear a systematic relationship to the three anatomical streams (P-B, P-I and M). The P-B stream is characterized by wave-length selectivity typically combined with a preference for low spatial frequency stimuli. Other types of selectivities - like orientation, direction and binocularity - have been reported to be low or absent. The P-I stream is characterized by high spatial resolution and selectivity for orientation, disparity and, to a lesser degree, also wavelength. It is the major source of orientation and disparity information reaching inferotemporal cortex from the high-spatial resolution parvocellular neurones of the retina. Combined together, these receptive field properties suggest that the parvocellular pathway is mainly involved in colour and high-resolution form perception.

The M stream has often been linked to motion analysis, mainly because of the high incidence of direction selectivity in the cortical cells of the stream, especially in area MT. The colour-blind M system has high temporal resolution, high contrast sensitivity and relatively low spatial resolution. The cortical M

cells also exhibit systematic selectivity for binocular disparity, suggesting that this system's other main concern is depth perception. One complexity is the prevalence in the M stream of selectivity for orientation. Livingstone and Hubel (1987) conclude that this indicates the system's ability to carry information about edge orientation, and suggest that this information is used for analysis of shape or for resolving ambiguities in stereo and motion matching. According to DeYoe and van Essen (1988), this property may be related to strategies for inferring 3-D trajectories. Gizzi et al. (1990) find it "interesting to consider the effects of this orientation selectivity on the way these neurones can signal visual motion". However, there are no real reasons why the simultaneous direction and orientation selectivity could not actually be a consequence of the motion coding. In fact, it is demonstrated in Chapter 6 of this thesis that cells can show concurrent direction and orientation selectivity if they are a part of a non-topographical map that codes velocity and form simultaneously.

Investigators have recently devised ways to block selectively each of the streams in rhesus monkeys to test their visual function - for a brief review, see Schiller and Logothetis (1990). The results of these studies are mainly consistent with the properties and functions of the streams presented above. They do, however, put more emphasis on the overlap of the functions. Parvocellular lesions severely impair colour discrimination and fine stereopsis. They also severely compromise contrast sensitivity and pattern vision at high spatial frequencies but not at low spatial frequencies. Magnocellular lesions have no discernible effect on colour vision, stereopsis or pattern vision. Most surprisingly, their effects on contrast sensitivity are only minor in spite of the fact that the cells in the M stream have been shown to be more contrast sensitive than those in the P stream. Magnocellular lesions affect most dramatically the perception of flicker and motion, especially at low contrasts. However, at low temporal frequencies and high contrasts the deficits are small, suggesting that the parvocellular system must be involved in these processes. To summarize, the results suggest that there is a range of vision served by both systems - the parvocellular system extends this range in the spatial and wavelength domains and the magnocellular system in the temporal domain.

2.2. Computational theories of early visual processing

The amount of information entering our visual system every single moment is huge. When the image is first captured by the photoreceptors in the eye, it consists of more that 100 million individual pointwise measurements of light intensity. The retinal circuitry uses centre-surround organization to decrease the correlation of neighbouring points, which effectively removes redundant data, but when the signals leave the retina, there are still about one million individual data values left. How is this large and unstructured set of measurements transformed into a more concise representation that is convenient for subsequent processing? This is the problem that most computational theories of early visual processing tackle. Another overload of information challenging the sensory system of every vision researcher is the vast quantity of psychophysical and neurophysiological information about visual processing. To absorb, analyze and compress this data, one needs theoretical frameworks and generalizations.

The problem of understanding visual perception in computational or information processing terms has been formulated most influentially by Marr and Poggio (1977; Marr, 1982). They argued that there are three distinct levels of analysis that are necessary for understanding a computational information processing task. At the top level is the computational theory, which analyzes what the problem is to be solved and why it is to be solved, and investigates the natural constraints imposed by the physical world on the solution of the problem. The next stage is to look for a particular algorithm or a detailed step-by-step procedure that transforms one representation of visual information into the next. At the lowest level is the implementation, a physical realization of the algorithm by some mechanism or hardware. Marr (1982) argued that the most efficient strategy for solving the computational problems of neuroscience was to move from the top level to the bottom; that is, one first formulates a theory, then derives an algorithm, and lastly designs a mechanism to implement the algorithm. The central claim of this approach is that it is logically possible to have a level of understanding of the visual processes without a knowledge of their underlying biological implementations. The extreme meaning of this claim - that computation and implementation can be separated and

considered entirely independently - has been generally criticized (Poggio, 1983; Morgan, 1985; Hildreth & Koch, 1987). These criticisms share the view that the nature of the hardware can profoundly influence the choice of algorithm. Gregory (1990) and Douglas and Martin (1991) have even argued that the whole concept of algorithm may be inappropriate in explaining the processes of an analogue device like the brain. However, despite its limitations, Marr's computational paradigm has made an important contribution to the study of vision by encouraging the generation of concrete and testable theories of human visual performance. The success of these theories also shows that there are situations where the visual processes can be described, at least to a certain extent, by sequential and parallel mathematical and logical operations, i.e. by algorithms.

This thesis has its main roots in two theories of early visual processing: the Fourier approach and the spatial primitives approach, both of which will be separately discussed below. The central idea of the Fourier approach is that the human visual system consists of separate channels, tuned to different spatial frequencies. The foundation of the spatial primitives approach rests on the theories of David Marr (1982). Its key idea is that the first stage in vision is the formation of a symbolic description of the image itself. A view common to many models based on the Fourier approach or the spatial primitives approach is that vision involves multiple spatial scales of analysis, processed in parallel.

There is also a third important class of computational theories that cannot be passed without mention - the neural network models, which are also referred to as adaptive networks, connectionist models, and parallel distributed processes. These models are neurally inspired models of brain and behaviour that try to mimic the structure and function of the processing brain. They show that a set of rather simple units can do very complicated things as a collective. By definition, a neural network is a parallel, distributed information-processing structure consisting of simple processing elements or neurones connected together with alterable strengths of connections. In vision research, these models have produced successful simulations of brightness perception (Cohen & Grossberg, 1984; Grossberg, 1987; Grossberg & Todorovic, 1988), form perception (Grossberg & Mingolla, 1985; Grossberg, 1987), fast perceptual learning in hyperacuity (Poggio, Fahle & Edelman, 1992), and motion perception (Grossberg & Rudd, 1992), among others.

However, most comparisons of network performance with psychophysical data have been only qualitative. The simulation of brain functions is both the strength and the weakness of the neural network approach. A successful reproduction of a given brain function using a neural network of a certain type, size and level of layers gives at least some indirect information how the brain may do the job. However, with increasing size and complexity of the networks, it becomes more and more difficult to analyze how they function. In other words, the networks may remain as black boxes to our understanding as the brain itself (Gregory, 1990). The computational modelling of this thesis does not directly apply the principles and techniques of neural networks, and due to the limited space they are not discussed separately, in spite of the fact that the discussion of the theoretical model for motion perception in Chapter 7 might have gained from this.

2.2.1. The Fourier approach

Much of the research in spatial vision during the last three decades has been influenced by the concepts of systems analysis and especially linear systems theory. Already in 1952 De Lange used some simple analogues borrowed from physics and engineering in his work on visual sensitivity to flickering light (De Lange, 1952). He showed empirically that the visual system behaves approximately linearly near flicker-detection thresholds. He pointed out that if one has measured contrast thresholds for sine-wave flicker at a sufficient number of flicker frequencies, it follows from linear systems theory that the threshold for any given waveform can be predicted by means of the inverse Fourier transform. Using this principle he was able to unify a large body of previous data on these thresholds. However, it was not until Campbell and Robson (1968) proposed that spatial patterns were analyzed into their spatial Fourier components and detected by specialized spatial frequency channels of the visual system that the application of spatial Fourier methods attracted wider attention. Their work has been one of the most influential initiatives in spatial vision research during the past few decades.

A basic principle in systems analysis is to compare the output signal with the input signal. If we represent the input to the system by variable x and the output by variable y and the system by an operator T, the system relation may be expressed as $y = Tx$. The

operator T transforms an input x to an output y. If we can express this statement in mathematical form, we have a mathematical model of the system. This model provides a functional identification of the system. In practice, with real physical systems the model is only valid for a restricted set of input conditions, nor does it reveal the internal structure of the system without *a priori* information about it (Regan, 1991).

Methods of visual psychophysics are inherently systems-analytical. These techniques analyse our visual system as a whole by presenting stimuli to the input and recording responses from the output. Using proper selections of stimuli, psychophysical experiments can sometimes break the black box of the visual system into smaller black boxes arranged sequentially or in parallel. Even after that, the functional description of the system often remains so complicated that most of the papers published in psychophysics have not tried to give any mathematical models of the system. However, starting from the papers of De Lange (1952) and Campbell and Robson (1968) there have also been good examples of the opposite: for example Wilson and Gelb (1984) and Watt and Morgan (1985). Chapter 4 of this thesis also provides an example of the use of mathematical modelling in the analysis of psychometric data.

The most important class of systems are the linear systems, even though real physical systems are never perfectly linear. A system $L\{\cdot\}$ is said to be linear if $L\{f(t) + g(t)\}$ equals $L\{f(t)\} + L\{g(t)\}$, that is, if the input is a sum of two functions, the output equals the sum of the outputs that would result if the inputs were fed to the system separately. In general a linear system is also assumed to be time-invariant, which means that any time shifts in the input waveform produce the same time shifts but no other changes in the output waveform. As a consequence of linearity and time-invariance, when the input to a linear system is a sine-wave, the output is a sine-wave with the same frequency as the input, though it may have a different amplitude and phase. If a system's output to a sine-wave input is not a sine-wave with the same frequency, the system is nonlinear. An important feature of any linear system is that if the output to a brief impulse or to a sufficient number of different input oscillation frequencies is known, the response to any given waveform can be computed.

The concept of frequency response characteristics has been very useful in the field of linear systems. This together with its tempting simplicity has also led to its use with nonlinear systems.

In principle this is acceptable if one remembers that in this case the model obtained is only descriptive; one cannot use it to predict the response to any given waveform. Often nonlinear systems behave linearly over some limited range of input variation. Our visual system is a system like this: it is overall nonlinear with many different kinds of nonlinearities like logarithmic compression, nonlinear inhibition, multiplication, division and rectification (Regan, 1991), but under some restricted conditions and over a limited range of inputs it or a part of it may behave linearly.

In visual psychophysics the typical stimulus or the input to the visual system is either a temporal or spatial variation of light intensity, or a combination of both. In terms of patterns, the stimulus can be a temporal, spatial or spatiotemporal pattern. We can describe a temporal pattern either in the time domain or in the frequency domain. These two alternative representations of the same data are related by the Fourier transform. Spatial patterns can be described in the space domain or in the spatial frequency domain, and Fourier methods can be applied to them equally well as to temporal patterns.

Fourier methods are the most commonly known group of methods in systems analysis and in signal analysis. Early in the 19th century Fourier showed that any waveform that is infinitely repeated along the time axis at a frequency of F can be represented as the sum of a series of sine waves whose frequencies are F, 2F, 3F, and so on. The F term is called the fundamental or the first harmonic component, the 2F term is called the second harmonic, and so on. The coefficients of the terms in the sum are obtained by calculating the Fourier transform of the original waveform. In general the Fourier transform is a complex-valued function of frequency; each term has both a real and an imaginary part. This has a physical meaning that at each frequency the Fourier transform gives the amplitude and phase of the sine at that frequency. In the real world we apply the Fourier transform to signals that are not infinitely repeated. These signals are called transient signals. Any real waveform is a transient waveform; even real-world sine waves are transients in that they do not extend to infinity. A transient waveform is described in the frequency domain as a continuous rather as a line spectrum. For example, a finite sine-wave train is equivalent to a band of frequencies rather than a single sharply-defined frequency. The shorter the transient, the

broader is the central lobe in the frequency domain description (Regan, 1991).

The transform from the time to the frequency domain is normally called the forward Fourier transform, and the one back from the frequency to the time domain the inverse Fourier transform. A fact often forgotten is that both the forward and the inverse transforms are global-to-local transforms. For example, in the inverse Fourier transform all the frequencies are taken into account in calculating amplitude at any one instant in time. Another fact that is even more often forgotten is that when the Fourier transform of a sampled spatial pattern is calculated using any of the fast Fourier transform algorithms now so common in digital signal processing, the algorithm assumes that the pattern is one cycle of a periodic pattern extending to infinity. Since in reality the pattern is usually restricted only to the sampled area, this leads to a distorted spatial-frequency spectrum - most commonly the portion of low spatial frequencies is underestimated. To minimize the distortion, one should extend the sampled area over which the transform is calculated well beyond the actual pattern.

It is often said (for example, Shapley and Lennie, 1985) that any real visual image can be represented as sum of sine-wave gratings of appropriate amplitudes and phases. Even though one could see this as the visual expression of the Fourier theorem which states that any waveform can be represented as a sum of sine waves, it is not entirely correct. A sine-wave grating is not a sine-wave distribution of light. Since there cannot be any negative light, a sine-wave grating is a sine-wave distribution plus a mean level that is always more than half the amplitude of the sine wave. Adding the gratings would progressively lower the contrast, thus rendering complicated patterns invisible (Regan, 1989). This is not to say that one could not use Fourier methods on real visual images. A real visual image can be represented as a sum of sine waves if only the sine-wave representation includes both positive and negative parts of the waves.

In their classic study on spatial frequency detection, Campbell and Robson (1968) determined the spatial frequency characteristics of the human visual system. They measured contrast thresholds for cosine gratings with different spatial frequencies, and presented their threshold data in a form of the contrast sensitivity function (CSF). Figure 2.1. presents the original CSF measured by Campbell and Robson. By assuming that the visual system was linear near

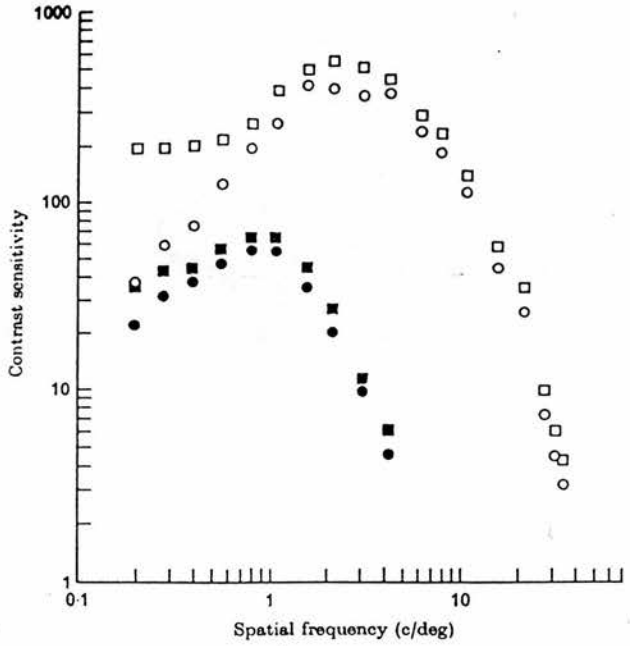


Fig. 2.1. Contrast sensitivity for sine-wave gratings (open and filled circles) compared with that for square-wave gratings (open and filled squares). Upper and lower pair of curves are for luminances of 500 and 0.05 cd/m^2, respectively. From (Campbell & Robson, 1968).

threshold, they further assumed that the contrast sensitivity function represented the Fourier transform of the spatial filter characterizing the visual system at threshold. To test their assumption they used this filter to predict contrast thresholds for square-wave gratings. Their measurements, however, did not confirm the predictions. Over a wide range of spatial frequencies the contrast threshold of a square-wave grating was determined only by the amplitude of the fundamental Fourier component of the grating. Campbell and Robson also noticed that to distinguish a square wave from its fundamental, the contrast of the square wave had to be raised to a level at which the higher harmonics reached their individual threshold. They concluded that the visual system at threshold could not be described by a single linear spatial filter, but, by a number of filters or channels each of which was tuned to a different frequency and had a significantly narrower bandwidth than the system as a whole. This idea gained vast support from both

psychophysical and neurophysiological studies in the following decades. Some of these studies are reviewed in Section 2.3.1.

The broad distribution of spatial frequency preferences and the narrow tuning of individual neurones has led many to propose that our visual system can be regarded as a Fourier analyser that generates a single frequency-domain representation from the retinal image. Shapley and Lennie (1985) provide a list of twenty such references, starting from Campbell and Robson (1968). They also provide a discussion about how the requirements for a rigorous Fourier analyzer are met in the visual system. These requirements include linearity, spatial delocalization, narrow bandwidth, encoding of amplitude and phase, and spatial homogeneity. Linearity of course is the basic requirement of the applicability of Fourier methods - however, the behaviour of many cortical cells is nonlinear. For instance, for many cells the line-weighting function cannot be predicted from the spatial frequency tuning curve (Movshon, Thompson & Tolhurst, 1978; Kulikowski & Bishop, 1981). In addition, line-weighting functions often have too few antagonistic subregions to be consistent with the spatial frequency responses under the linearity assumption. In other words, they are too localized in space - contrary to spatial delocalization requirement. Another problem with the global Fourier mechanism is phase encoding. It is known that the appearance of many natural images (such as faces) is specified to a much greater degree by their phase than by their amplitude spectra (Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982). However, no satisfactory explanation of how the phase information might be preserved has been offered (Regan, 1991). Spatial homogeneity requirement is violated by the fact that the receptive field sizes of retinal ganglion cells and cortical cells are not uniform across the visual field (Cleland, Harding & Tulunay-Keesey, 1979; Rovamo & Virsu, 1979, among others). Thus, most of the requirements for Fourier analysis are not met by the visual system. An important qualification of the Fourier representation approach was presented by Robson (1975). He proposed that each small region of the retinal image is analyzed into a limited number of frequency bands. This idea of patchwise Fourier analysis fits much better to the properties of the visual system than the global Fourier analyzer. It also diminishes the problem with the phase information - Morgan, Ross and Hayes (1991) showed that with small patchsizes the phase information loses its dominance in the appearance of the image to the amplitude information.

Several models of localized Fourier-like analysis have been proposed to account for spatial pattern discriminations and hyperacuity (Wilson & Gelb, 1984; Klein & Levi, 1985; Nielsen, Watson & Ahumada, 1985, among others). In the model of Wilson and Gelb, which they call a line element model for spatial pattern discrimination, there are linear spatial filters stationed at every retinal point, providing the localization of the signal. At each position, there are several filters with different preferential spatial frequencies and orientations, and with moderately narrow bandwidths. This accomplishes the Fourier-like analysis. The result of each filter is passed through a contrast nonlinearity, and these responses, with added noise, then form a representation in multi-dimensional mechanism response space. Each spatial pattern is represented as a point in this space, and pattern discrimination is based on the distance between these representations. Based on the results of Wilson, McFarlane and Phillips (1983), Wilson and Gelb used six different spatial frequency tuned mechanisms in their model. Results of model computations agree reasonably well with data from many spatial pattern discrimination experiments by several investigators (Wilson, 1991). The models of Nielsen et al. (1985) and Klein and Levi (1985) differ from the line element approach in the use of a larger number of different mechanism sizes with both odd and even symmetries, but the basic principle is similar.

2.2.2. The spatial primitives approach

Models of localized Fourier-like analysis assume that hyperacuity limits depend on sensitivity and size of spatial filters but they do not make any assumptions about the representation of space in the visual system. An alternative view is that hyperacuity actually reflects the scale or accuracy of the internal representation of visual space. This view originates from the scheme for spatial representation of Marr (1982). He argued that the initial spatial distribution of intensity values across the retina is too large and unstructured to allow efficient manipulation at subsequent processing stages. He suggested that the first stage in vision is the formation of a symbolic two-dimensional spatial description of the retinal image; this description was termed in Marr's theory the primal sketch. The primal sketch contains primitive descriptors or tokens that are derived by identifying important features in the retinal image: edges, bars, blobs, texture boundaries etc. The

later stages of computations are implemented as manipulations of the primitive descriptors.

The important features in the retinal image are always accompanied by significant intensity changes at a number of spatial scales. Since edges can be defined as locations of significant intensity changes, they are a good candidate for the basic spatial primitive. In fact, a representation of the edges in the image has proved to be a useful first step in many applications in the fields of computer vision and image processing. Early perceptual studies by Cornsweet and others (see Cornsweet, 1970) have also shown that our perception of an image is strongly influenced by the presence of sharp intensity changes or edges. Marr and Hildreth (1980) suggested that the importance of edge analysis may be related to the fact that in our visual system, the incoming image is initially filtered by the difference of Gaussian or DOG-shaped receptive fields of the retinal ganglion cells. They argued that zero crossings in the DOG-filtered image correspond closely to intensity edges in the original image.

To see the relation between zero crossings and edges, one must first consider the problem of detecting edges in a one-dimensional signal (for a concise review, see Ullman, 1986). Since an edge is a significant change in the signal, it can be located by detecting local maxima or minima in the signal's derivative, or by detecting zero crossings in the second derivative - zero crossings in the second derivative correspond to peaks and troughs in the first. Differentiation of a signal, however, tends to amplify the high-frequency noise components in the signal. This problem can be overcome by attenuating the high-frequency components prior to the differentiation operation. Marr and Hildreth (1980) showed that this can be done optimally by using a smoothing function that has a Gaussian shape. Thus, one procedure for edge detection is to perform the following steps: (1) smooth the signal using a Gaussian filter, (2) perform a second derivative operation, and (3) locate the zero crossings. However, it can be shown that Gaussian smoothing followed by a second derivative operation is equivalent to a single step in which the signal passes through a second derivative of a Gaussian. Finally, Marr and Hildreth observed that the second derivative of a Gaussian is almost identical in shape to a difference of Gaussians (DOG), and concluded that DOG-like receptive fields can be used for edge detection.

Marr and Hildreth's (1980) proposal that the fundamental spatial primitive is the zero crossing in the DOG-filtered retinal stimulus gained support from Watt and Morgan (1983a). In a vernier acuity task, their data were well fitted by a model that located zero crossings in the signal filtered by a single-size filter - Marr and Hildreth had claimed that zero crossings provide a scale-independent representation of edges. However, there are many cases where this is not true: for instance, where several edges are near neighbours, or where information actually exists at different spatial scales (Watt & Morgan, 1983b; 1984). Watt and Morgan (1983b) also showed that observers can distinguish different amounts of edge blur, and discriminate different types of blurring function, indicating that zero crossings cannot be the only form of information extracted from edges.

The problems with the zero-crossings model led Watt and Morgan (1983b; 1984; 1985) to revise their single-size filter model. The result was the MIRAGE model (Multiple Independent filters, half-wave Rectified, Averaged and Gated for Extraction of the primitive code), the most elaborate example of the spatial primitives approach. MIRAGE starts with four 1.7 octave filters spaced at one octave intervals from 3.4 - 27.2 c/deg. The filters are presumed to have balanced positive and negative weighted inputs in a circularly symmetric centre-surround arrangement, with the Laplacian of a Gaussian as a standard form. Each filter is independently convolved with the retinal image, and the convolutions are then separated into their positive and negative parts via half-wave rectification. All the positive responses or peaks from all filters are added together to produce a net positive signal, and the same is done quite separately to the negative responses or troughs. Thus, unlike many other models, MIRAGE does not produce separate representations at different spatial scales. Each signal, the net positive and the net negative one, is separately broken down into a sequence of zero-bounded response distributions and regions of zero-response. To characterize zero-bounded distributions, Watt and Morgan proposed the use of the statistical central moments of the activity distributions. These are the centroid, which localizes the distribution; the mass or area, which provides a measure of the response magnitude; and the standard deviation, which gives an estimate of the spatial extent of the activity and of the accuracy of localization. A set of rules is then used to determine whether an edge or line caused the response.

MIRAGE has been especially successful in predicting data on positional acuity and edge blur discrimination (summarized by Watt, 1988). One of the findings leading to the development of MIRAGE was that the distance between the peak and the valley in the Laplacian of a Gaussian filtered edge encodes the extent of edge blur in a way very similar to the performance of the human observer (Watt & Morgan, 1983b). The main criticism presented against MIRAGE is that it does not appear to make significant use of the multiple spatial filter sizes that are almost unanimously believed to exist in the visual system (Wilson, 1991). Recently, Kingdom and Moulden (1992) have presented a model for brightness coding that is in many respects very similar to MIRAGE, but in which the responses of the spatial filters are interpreted before they are combined.

2.3. Spatial and temporal properties of vision

There is much psychophysical and neurophysiological evidence that the apparent subunits of the visual system - channels, mechanisms, filters, detectors, neurones etc. - are specifically sensitive to particular values of certain primitive attributes of visual stimuli. These attributes include for instance spatial frequency, temporal frequency, orientation, direction of movement, binocular disparity and colour. This section discusses the sensitivities of the basic mechanisms related to the perception of moving objects.

2.3.1. Spatial frequency channels

In the classic study on spatial frequency detection Campbell and Robson (1968) used cosine gratings to measure the contrast sensitivity function of the human visual system (see Figure 2.1). They found that the sensitivity shows bandpass characteristics, peaking at a spatial frequency of about 3 cycles/deg. This contrast sensitivity function could not be explained by a single linear spatial filter. Instead, Campbell and Robson proposed that the visual system at threshold must be composed of a number of independent filters or channels, each of which was tuned to a different spatial frequency and had a significantly narrower bandwidth than the system as a whole.

The principle that the visual system conveys information about spatial frequencies in tuned channels gained early support from Blakemore and Campbell (1969) and Graham and Nachmias (1971). This led to an increasing number of studies concentrating on defining the characteristics of the individual filters. A review and a thorough summary of the results of these studies is given by Wilson (1991). Filter sensitivities were mainly estimated using three different psychophysical paradigms: spatial frequency adaptation, subthreshold summation and pattern masking. The spatial frequency adaptation paradigm was introduced by Blakemore and Campbell. They showed that a several minutes adaptation to a high-contrast sinusoidal grating (7.1 c/deg) selectively elevates the contrast threshold for test gratings of similar spatial frequency and orientation. The degree of spatial tuning is usually quantified in terms of the bandwidth, which is the ratio of the higher to the lower frequency at which the sensitivity is declined to one half the peak sensitivity. For the

tuning function obtained by Blakemore and Campbell, the bandwidth was about one and a half octave. In the literature of spatial vision the meaning of octave is analogous to its meaning in music; an octave is an interval between frequencies one of which is twice the other, and not anything eight-fold as the word might hint.

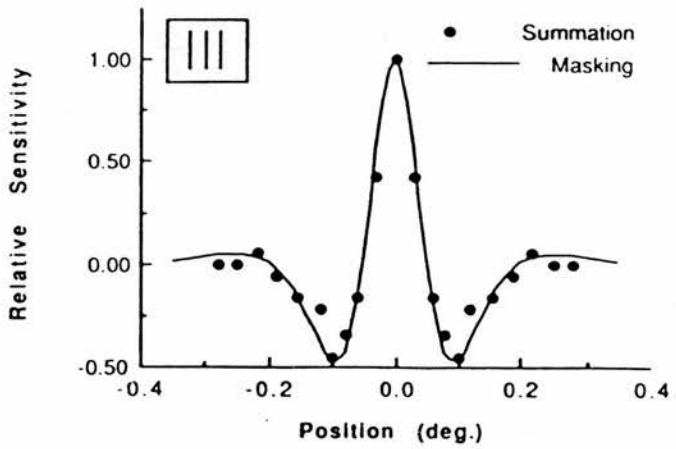


Fig. 2.2 Data from an experiment measuring subthreshold summation between a centre line and two flanking lines. A central excitatory zone is surrounded on either side by an inhibitory trough (Wilson, 1978). The solid curve is the inverse Fourier transform of the curve fit to the masking data corresponding to a spatial frequency mechanism with a peak frequency of 4 c/deg (Wilson, McFarlane & Phillips, 1983). From Wilson (1991).

A bandwidth of 1.5 octaves for spatial frequency tuned mechanisms has been later confirmed by many studies using the subthreshold summation paradigm (King-Smith & Kulikowski, 1975; Stromeyer & Klein, 1975; Wilson, 1978; Bergen, Wilson & Cowan, 1979; Robson & Graham, 1981, among others). This technique measures the threshold for a test pattern in the presence of a range of patterns at fixed subthreshold contrasts, and uses the data to estimate spatial frequency characteristics of the underlying mechanisms. Fig. 2.2 shows results from an experiment where thresholds for a central line were measured as a function of the position of a pair of subthreshold flanking lines (Wilson, 1978). When the central line and the flanking lines are superimposed, or when the separation between them is small (less than about 0.05° or 3′ of arc), spatial summation enhances detection of the central line. When the separation is increased beyond 0.05°, the threshold increases until

0.2°, indicating lateral inhibition. Beyond 0.2° the flanking lines have no effect. The Fourier transform of this line-weighting function produces a contrast sensitivity function with a bandwidth of 1.5 octaves.

Masking is the third technique that has been used to study the tuning properties of visual mechanisms (Legge & Foley, 1980; Wilson, McFarlane & Phillips, 1983). This technique is based on the interaction between a clearly visible masking pattern and a superimposed test pattern. In the typical paradigm, the properties of the test pattern remain fixed while the mask varies along the dimension of interest. The reduction of sensitivity as a function of the mask variable is measured. Wilson et al. (1983) used spatially localized test patterns with a 1.0 octave bandwidth Fourier spectrum and cosine grating masks oriented at a slight angle relative to the test pattern. As masking had been shown to be a compressive nonlinear function of mask contrast (Nachmias & Sansbury, 1974; Legge & Foley, 1980), Wilson and colleagues measured this nonlinearity and corrected their data for it. Their final results for foveal vision were consistent with the operation of six spatial frequency tuned mechanisms with peak frequencies ranging at about one octave intervals from 0.8 - 16 c/deg. The bandwidths of the sensitivity curves were found to decrease from about 2.2 octaves at 0.8 c/deg to about 1.3 octaves at 16.0 c/deg.

Neurophysiological studies on spatial frequency tuning of cortical neurones have provided results that are in reasonable accord with the psychophysical results. Many cortical neurones in the cat are tuned to spatial frequency (Campbell, Cooper & Enroth-Cugell, 1969; Maffei & Fiorentini, 1973), and this tuning is narrow for both simple and complex cells (Ikeda & Wright, 1975; Tolhurst & Movshon, 1975; Movshon, Thompson & Tolhurst, 1978). Similar results have also been found for the monkey's striate cortex cells (De Valois, Albrecht & Thorell, 1982). The average bandwidth of visual cortical cells in both cats and monkeys is around 1.5 octaves, but the range of bandwidths is very large. For most cells, the bandwidth is in the range from 1.0 - 1.5 octaves, but bandwidths below 1.0 octave or above 2.0 octaves are not uncommon. In the region of the cat's cortex representing the area centralis, spatial frequency preferences vary from 0.3 up to 3.0 c/deg (Movshon, Thompson & Tolhurst, 1978). For the monkey visual cortex representing the fovea, the range is somewhat larger with a median near 3.0 c/deg (De Valois, Albrecht & Thorell, 1982).

2.3.2. Sensitivity to temporal variations

The temporal properties of the human visual system were first studied by measuring critical flicker fusion (CFF) limits. This value is the maximum rate at which flicker can be resolved. Over a considerable luminance range, the CFF obeys the Ferry-Porter law (Ferry, 1892; Porter, 1902), which states that the CFF varies approximately proportionally to the logarithm of luminance. At about 1000 Td - a typical illuminance level for television set - the CFF is just under 50 Hz (Burr, 1991).

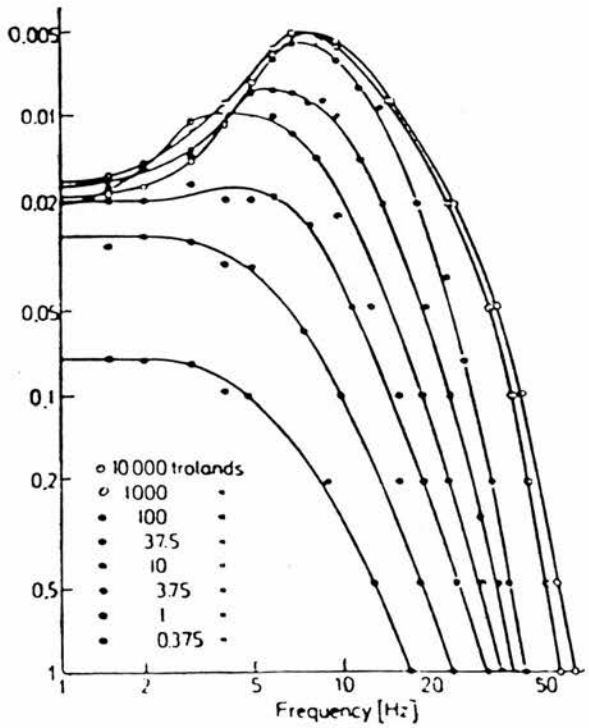


Fig. 2.3 Contrast thresholds for detecting the flicker of a 2° patch of light as a function of the temporal frequency of modulation, for different levels of illumination (from De Lange, 1958).

De Lange (1952; 1958) measured the minimum amplitude of sinusoidal modulation of a 2° patch of light required for detection of flicker, over a range of temporal frequencies and illumination levels, thus producing temporal contrast sensitivity functions (CSF). Fig. 2.3 shows examples of De Lange's results. At high

illumination levels, sensitivity is maximal at about 8 Hz, and fall with lower or higher temporal frequencies. At lower illumination levels, the CSFs shift downwards and to the left, implying lower absolute sensitivity and lower temporal resolution or CFF. The results show that the whole temporal response varies with luminance. Kelly (1961) made similar measurements using a much larger (65°) flickering field. This shifted the peak sensitivity to around 20 Hz at high luminances. Later Kelly (1972) replotted this data and De Lange's data as absolute amplitude thresholds, and noticed that at all luminance levels, the curves for both small and large patches tend to fall within a single envelope. This shows that near the limit of critical flicker fusion, resolution is limited by absolute modulation, irrespective of adaptation level. Thus, for temporal frequencies near the CFF, the visual system behaves linearly.

Robson (1966) measured contrast sensitivities as a function of both spatial and temporal frequencies. The stimuli were sinusoidal gratings caused to reverse in contrast sinusoidally. Gratings with this kind of temporal modulation are often referred to as counterphase gratings. Robson found a reciprocal relationship between spatial and temporal tuning. For low spatial frequencies (0.5 c/deg in Robson's data), the temporal contrast sensitivity function is bandpass, with a peak at about 6 Hz. For moderate and high spatial frequencies (4, 16 and 22 c/deg), the temporal CSF is lowpass. The spatial CSF is bandpass for low temporal frequencies, and lowpass for high frequencies. Low frequency attenuation occurs only when both temporal and spatial frequencies are low.

Human eyes are in constant motion produced by the oculomotor system even when one tries to maintain steady fixation. These movements are known to influence the detection of stationary gratings at low temporal frequencies (Kulikowski, 1971). To avoid this effect, Kelly (1979) measured the spatiotemporal contrast-sensitivity threshold surface using stimuli stabilized on the retina. He found that when stabilized stationary and moving gratings are superimposed, it is only the moving grating which controls the threshold - the stationary grating fades away. Kelly proposed that the visual system responds only to a certain range of stimulus velocities and that the eye movements must provide the image movement for otherwise stationary patterns on the retina. To support this idea, he showed that the spatial contrast sensitivity function for unstabilized stationary gratings can be matched quite

well by moving the stabilized retinal image at the subject's spontaneous eye-movement drift rate (about 0.1 - 0.2 deg/s).

The overall spatiotemporal contrast-sensitivity threshold surface obtained by Kelly (1979) confirmed the results of Robson (1966). The spatial CSF is bandpass for low temporal frequencies and lowpass for high temporal frequencies, and the same is reciprocally true for the temporal CSF. Robson had originally proposed that this behaviour could result from the centre-surround organisation of retinal ganglion cells, and later Burbeck and Kelly (1980) developed this idea further. They assumed that as the centre and the surround of the receptive field of retinal ganglion cells have different spatial-frequency characteristics (Enroth-Cugell & Robson, 1966), they may also have different temporal characteristics. Burbeck and Kelly showed that the spatiotemporal threshold surface by Kelly (1979) can be modelled as the difference between two other spatiotemporal surfaces, both of which have a simple lowpass form but one of which (centre) is broader than the other (surround) in both spatial- and temporal-frequency bandwidth. This is simply a spatiotemporal analogue to the explanation for the bandpass behaviour of the difference-of-Gaussians type spatial filters or receptive fields. Several neurophysiological studies have found different temporal characteristics for the receptive field centre and surround (Barlow & Levick, 1969a; Barlow & Levick, 1969b; Derrington & Lennie, 1984; Fleet, Hallett & Jepson, 1985), though in some studies the temporal aspects of centre-surround interactions could be better explained by the existence of a small increase in latency of the surround response compared with that of the centre (Derrington & Lennie, 1982; Enroth-Cugell, Robson, Schweitzer-Tong & Watson, 1983).

Temporal frequency channels

Evidence that the visual system contains mechanisms tuned to different spatial frequencies was reported in the classic study of spatial frequency detection by Campbell and Robson (1968). Corresponding mechanisms for the temporal domain, however, proved to be more difficult to find. The first clues came from the finding that different thresholds are found for flicker and form detection of drifting gratings (Van Ness, Koenderink, Nas & Bouman, 1967) or of a flickering bar (Keesey, 1971). Kulikowski and Tolhurst (1973) used counterphase modulated gratings to measure temporal modulation

transfer functions for both flicker and pattern detection. The former showed a spatial lowpass and a temporal bandpass response function and the latter a higher spatial bandpass and a temporal lowpass response function. Kulikowski and Tolhurst suggested that two independent populations of mechanisms underlie the detection of grating patterns of low and high spatial frequency, and that these mechanisms have different temporal properties: transient mechanisms or detectors prefer low spatial and high temporal frequencies, while sustained mechanisms prefer high spatial and low temporal frequencies.

Watson and Robson (1981) provided evidence for the existence of two broadly tuned mechanisms in temporal frequency. For gratings flickering at different temporal frequencies, they observed that two temporal frequencies can be perfectly discriminated only if one is higher than about 8 Hz and the other less than 2 Hz. They concluded that the temporal frequency dimension is served by two discrete mechanisms: one selective for temporal frequencies from about 6 to 10 Hz for moving targets, and the other for stationary or slow moving targets. Using both at-threshold and suprathreshold temporal discrimination tasks for drifting gratings, Thompson (1983) confirmed the main finding of Watson and Robson.

Hess and Plant (1985) measured suprathreshold temporal frequency discrimination over a wide range of spatial and temporal frequencies. They found a single minimum (at about 4 Hz) in the discrimination function at 2 c/deg but also a second minimum (around 16 Hz) at 0.2 c/deg, providing evidence for a third mechanism with a poor spatial resolution and a temporal passband peaking above 10 Hz. This finding is consistent with the results of Mandler (1984), who measured temporal frequency discrimination using a 1° flickering spot on a broad surround. His results pointed to three mechanisms, one of which was lowpass and the other two broad bandpass, with peak sensitivities at about 5 and 15 Hz.

Neurophysiological studies on temporal frequency tuning in cats have shown that the attenuation at high temporal frequencies is likely to result from an integration process localized between the lateral geniculate nucleus and the cortex. Derrington and Lennie (1984) showed that ganglion cells and cells in the LGN of the cat will respond at temporal frequencies over 100 Hz. Cells in areas 17 and 18 of cat visual cortex, however, show peak responses at 4-5 Hz, and the response is considerable attenuated at 25-30 Hz (Movshon, Thompson & Tolhurst, 1978). Human steady-state flicker visual

evoked potentials measured from the occipital scalp, on the other hand, have revealed three subsystems broadly tuned to about 10, 16 and 40-50 Hz, as reviewed by Regan (1989). Since these measurement have been done using unpatterned light, one can speculate that the two lower mechanisms may be related to the psychophysical temporal frequency channels of Hess and Plant (1985). The 40-50 Hz mechanism, however, has no corresponding mechanism in the psychophysical data described above.

Temporal impulse response

Temporal weighting functions can be used to describe the properties of temporal frequency channels in the temporal domain. These functions are analogous to receptive fields or point- or line-weighting functions that describe spatial frequency channels or filters in the spatial domain. The most common temporal weighting function is the temporal impulse response, which - like the impulse response in general - refers to the response to a very brief pulse.

The temporal impulse responses presented in the literature are typically biphasic, having an excitatory phase followed by a smaller inhibitory phase (Watson, 1982). A representative example is shown in Fig. 2.4. This curve was obtained by McKee and Taylor (1984) by fitting their foveal flicker threshold data to a model proposed by Watson (1982). This model is basically a difference of two gamma functions, and similar models have been successfully used to fit both psychophysical (Levinson, 1968; Bergen & Wilson, 1985) and neurophysiological data (Baylor, Hodgkin & Lamb, 1974). The temporal extent of the impulse response depends strongly on the adaptation level - the impulse response becomes shorter when the retinal illuminance increases (Kelly, 1971; see also a review by Ikeda, 1986). The presence of inhibition, on the other hand, depends on the size of the stimulus. Meijer, van der Wildt and van den Brink (1978) studied the twin-flash response as a function of flash diameter, and found that the stimulus size must be larger than about 11 arc min for the inhibitory phase to appear. Watson and Nachmias (1977) had found the same effect with pairs of brief, temporally separated gratings of the same spatial frequency. For low spatial frequencies, they found a range of separation that resulted in inhibition between pairs. For higher spatial frequencies, this range was absent or nearly absent.

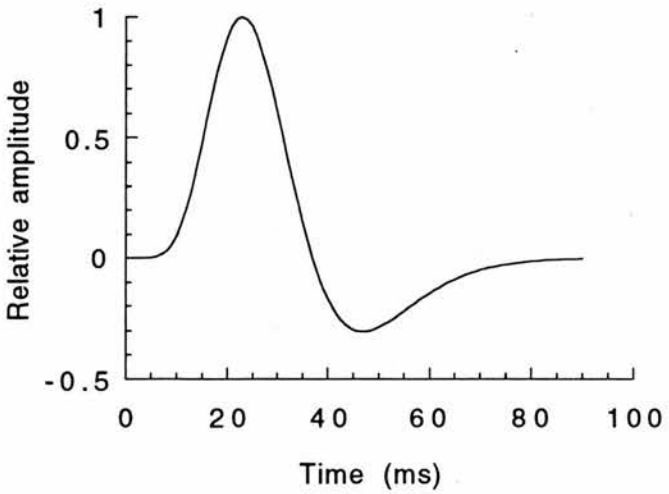


Fig. 2.4. A temporal impulse response function for foveal vision with an adapting background luminance of 50 cd/m^2 (redrawn from McKee and Taylor, 1985).

If the impulse response of a linear system is known, then the response to any input waveform can be predicted. Unfortunately, the temporal characteristics of our visual system are nonlinear even at threshold. Kelly and Savoie (1978) reported a small but systematic asymmetry in the visual response to flashed fields. This essential nonlinearity (i.e. one that cannot be linearized even for small signals) had a greater gain for the negative parts of the filtered waveform that for the positive parts. Another important nonlinearity in determining temporal thresholds appears to be temporal probability summation (Watson & Nachmias, 1977; Watson, 1979). Bergen and Wilson (1985) showed that at low spatial frequencies one needs both of these nonlinearities to predict accurately sinusoidal flicker sensitivities from thresholds for a series of three temporal pulses, while at high spatial frequencies temporal probability summation is the only nonlinear factor required for accurate prediction. Due to these nonlinearities, the psychophysically measured temporal impulse response functions are only descriptive functions, and not real impulse response functions in the systems-analytical meaning of the term.

2.3.3. Sensitivity to motion

With visual motion, there has always been some doubt whether it actually is a fundamental visual dimension or whether it is derived from space and time, as motion is in physics. At the first glance, any logical person might think that it must be a derived quality since motion does not exist without change in position with time. However, visual motion is a perception, not a simple copy of physical motion. A lot of evidence supports the view that visual motion is indeed served by its own separate process (Nakayama, 1985). The oldest demonstration (Purkinje, 1825) is the motion aftereffect or the "water-fall" illusion: stationary objects are seen as moving in a direction opposite to that of a previously viewed motion. Another good example is direction specific adaptation (e.g. Tolhurst, 1973).

If visual motion is a fundamental visual dimension, then one can raise a question about the independence of the temporal mechanisms of the visual system. Many of the experiments described above in connection with temporal selectivity used counterphase gratings. Levinson and Sekuler (1975) pointed out that a counterphase grating is physically identical to the sum of two moving gratings, each of half amplitude, travelling in opposite directions. They and Watson et al. (1980) showed that thresholds for detecting counterphase gratings were almost twice as high as those for detecting drifting gratings. This is what would happen if counterphase gratings were detected by directionally tuned mechanisms that respond to the drifting components of the counterphase grating, and there was no probability summation between the mechanisms tuned to opposite directions. Watson et al. also showed that at low temporal and high spatial frequencies, sensitivity to counterphase gratings approaches that of drifting gratings, indicating detection by directionally non-selective mechanisms at those frequencies. The results suggest that flicker and motion perception are related, and may be mediated by the same mechanisms, at least at low spatial and high temporal frequencies.

In the previous sections, temporal properties of the visual system were discussed mainly on the basis of experiments using flickering targets. In the following, sensitivities to moving objects will be discussed by referring to results from experiments that have mainly used moving targets. As shown above, this distinction may be highly artificial, and thus the visual

characteristics described in these two sections should be considered with the possibility in mind that they or a part of them are properties of the same set of mechanisms. It should also be implicit in the discussion so far that some kind of partial inseparability actually holds for all the relations between spatial and temporal properties of the visual system.

Drifting sinusoidal gratings have been one of the most-used stimuli in measuring the properties of motion perception. Tolhurst (1973) showed that adapting to a drifting sinusoidal grating causes threshold elevation that is both direction and spatial frequency specific. Adaptation failed to reveal channels at low spatial frequencies when stationary gratings were used. Tolhurst regarded this as evidence for movement-sensitive channels at low spatial frequencies. He suggested that these movement-sensitive or transient channels alone convey information leading to the sensation of movement, while movement-independent or sustained channels at higher spatial frequencies give no information about temporal changes but about the structure of the stimulus, whether moving or stationary. Thus, Tolhurst's proposal clearly separates the perception of form and motion of a moving object.

Kelly (1979) showed that if contrast sensitivity functions to drifting gratings are measured at constant velocity, a family of curves differing in optimum spatial frequency are produced. Fourier transforms of these curves yield a range of receptive field or filter sizes in the space domain, the larger filters being tuned to higher velocities. His results were confirmed by Burr and Ross (1982) who obtained the curves shown in Fig. 2.5 using motion thresholds, the minimum contrast required to see the direction of motion. Spatial contrast-sensitivity functions for five different drift speeds are summarized in Fig. 2.5(a). The curves all have the same shape and size; drift speed only changes the position of the curves on the spatial frequency axis: the higher the velocity, the lower the optimum spatial frequency. In other words, motion does not lower absolute sensitivity nor diminish the visual passband, but instead slides the resolvable spatial frequency window along the spatial frequency scale. Fig. 2.5(b) plots the same data as a function of temporal frequency rather than spatial frequency. The curves at all velocities (except 1 deg/s) are very similar, peaking at about 10 Hz, and extending to frequencies of about 50 Hz. Thus, the upper limits for motion perception are not set by velocity, but by temporal resolution. The curve for the slowest velocity, 1

deg/s, is slightly different from others, consistent with the idea of different mechanisms for slow and fast motion (Kulikowski & Tolhurst, 1973; Tolhurst, 1973; Watson & Robson, 1981). To confirm that their results also extend to single objects, Burr and Ross showed that contrast-sensitivity behaviour for a single cycle of sinusoidal grating in motion is similar to that of gratings.

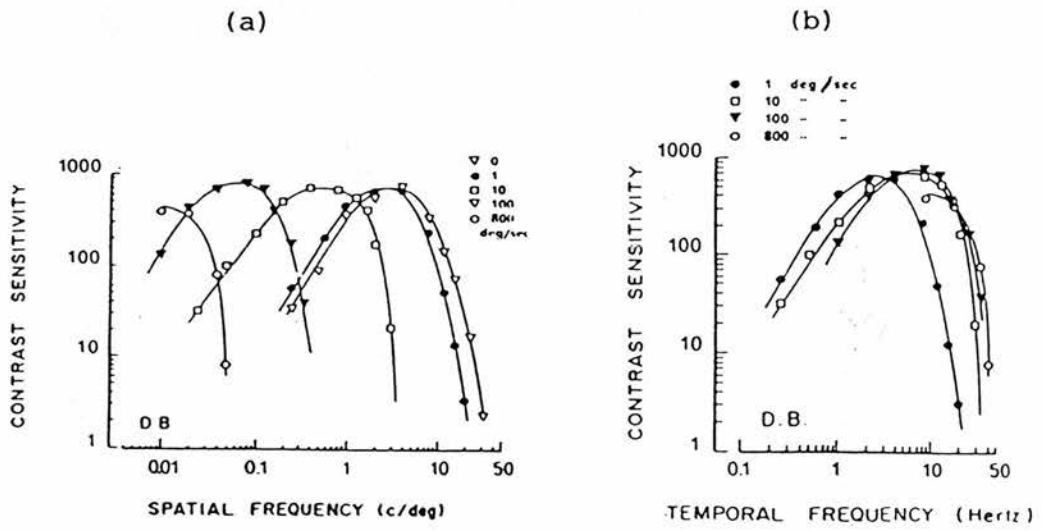


Fig. 2.5 Contrast-sensitivity curves for detecting the direction of drift of a sinusoidal grating for five different velocities (a) as a function of spatial frequency, and (b) as a function of temporal frequency (from Burr and Ross, 1982).

Anderson and Burr (1985) studied motion mechanisms by measuring sensitivity to moving gratings in the presence of masks. For every spatial frequency of the test (ranging from 0.07 to 20 c/deg), masking was maximal when the test and mask frequencies coincided, implying that there exist detectors of limited bandwidth tuned to that spatial frequency. The bandwidth of the spatial frequency masking functions decreased monotonically with spatial frequency from three octaves at 0.06 c/deg to one octave at 30 c/deg. This decrease agrees with the findings of earlier psychophysical studies (e.g. Wilson et al., 1983). Temporal frequency masking followed a quite different pattern from spatial frequency masking. Fig. 2.6 shows masking functions for a representative low temporal frequency (0.7 Hz) and high temporal frequency (10 Hz), for three spatial

frequencies. The high temporal frequency function is bandpass with a peak sensitivity from 7 to 13 Hz (depending on spatial frequency), but the low temporal frequency function is lowpass, sensitive up to about 13 Hz. At high spatial frequencies, the lowpass function is higher or more sensitive than the bandpass function, at low spatial frequencies, the reverse holds. The results are consistent with the previous ideas of two groups of temporally tuned visual detectors, transient and sustained (e.g. Tolhurst, 1973).

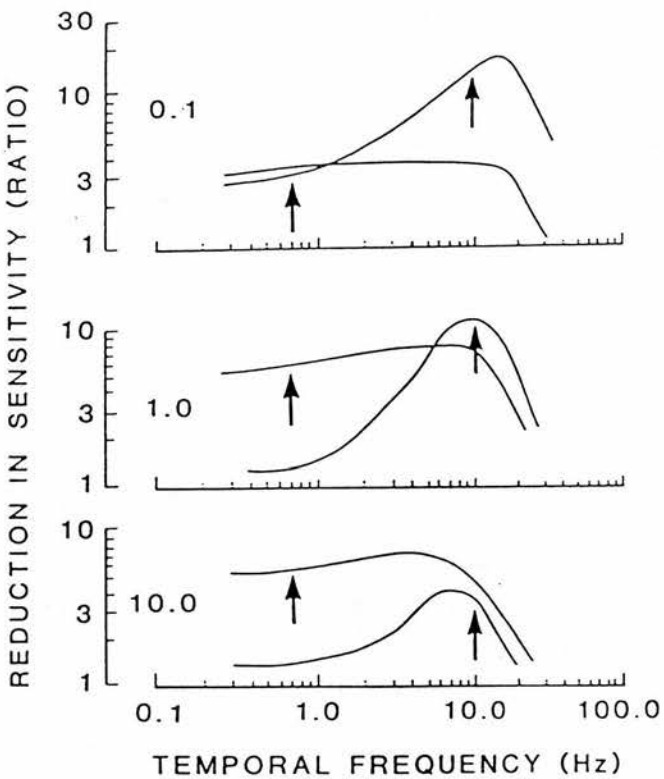
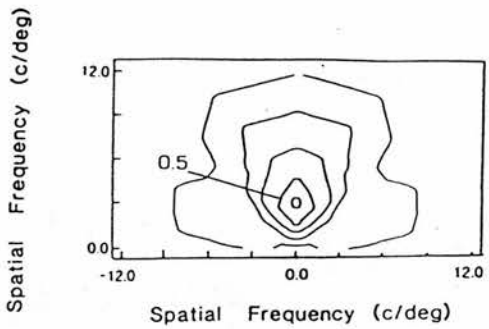


Fig. 2.6 Masking as function of the temporal frequency of the masking stimulus. The test was a vertical sine-wave grating of 0.1 (top), 1.0 (middle) or 10.0 c/deg (bottom). It was drifted at 0.7 or 10.0 Hz (arrows). The mask was one-dimensional random noise reversing in contrast at constant temporal frequency, indicated by the abscissa. Two basic types of temporal functions, lowpass and bandpass, are revealed (from Anderson and Burr, 1985).

(a)



(b)

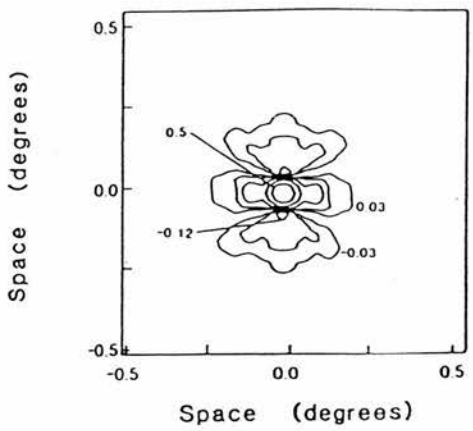


Fig. 2.7 (a) The two-dimensional spatial frequency tuning function produced with jittering mask gratings for detectors responding to a stimulus grating of 3.0 c/deg drifting at 8 Hz. (b) The contour map of the psychophysical receptive field calculated by the inverse Fourier transform of the tuning surface in (a) (from Anderson, Burr and Morrone, 1991).

Anderson, Burr and Morrone (1991) measured thresholds for detecting the direction of motion of drifting vertical gratings in the presence of masks that varied in both spatial frequency and orientation. The masking data were scaled appropriately to give an estimate of the two-dimensional spatial-frequency tuning surface of cortical detector units in human vision. Fig. 2.7 (a) shows an example of this tuning surface for detectors tuned to 3 c/deg. With the assumption of small signal linearity and zero phase, the tuning surfaces were inverse Fourier transformed to give an indication of the size and structure of the psychophysical receptive fields of detector units. Fig. 2.7 (b) shows the transform of the tuning

surface of Fig. 2.7(a). The receptive field is made up of several adjacent opponent regions and is oriented in two-dimensional space, giving it a selectivity for both spatial frequency and orientation. The receptive field size varies inversely with spatial frequency, being largest for units tuned to low spatial frequencies. At all spatial scales, the width of the field is roughly the same as the length.

Anderson and Burr (1987; 1991) have also studied the receptive fields of motion detectors by more direct means using a summation technique. They measured contrast thresholds for small patches of drifting gratings, curtailed either in length or in width. Sensitivity to the gratings increased as a function of their length or width, first linearly with a slope of one on a log-log plot and then more gradually with a slope of about 0.29. The slope of one was taken as being indicative of total or physiological summation and the lesser increase of probability summation. The extent of physiological summation is a logical estimate of the spatial extent of a receptive field. The estimates from the summation studies (Anderson and Burr, 1987; Anderson and Burr, 1991) agree closely with the estimates obtained from Fourier transforms of masking data (Anderson and Burr, 1989; Anderson, Burr and Morrone, 1991), suggesting that receptive fields of human motion detectors are as long as they are wide, and that the size varies inversely with spatial frequency, being about 6 deg at 0.03 c/deg and less than 0.1 deg at 10 c/deg.

Motion combines space and time, and thus the spatial organization of receptive fields of motion detectors describes only one part of their properties. Burr, Ross and Morrone (1986) used a masking technique to measure the joint spatiotemporal tuning functions of motion detectors, and assuming linearity at threshold, they calculated the spatiotemporal receptive fields of the detectors by inverse Fourier transform of the data. The fields obtained comprise alternating ridges of opposite polarity, elongated in space-time along the preferred velocity axis of the detector. Burr et al. suggest that this organization explains how motion detectors analyze form and motion concurrently. This principle is opposite to the idea proposed by Tolhurst (1973) which separates the perception of form and motion of a moving object. The concept of spatiotemporal receptive fields is one of the main explanations of the blur-free perception of moving objects, and thus one of the main topics of this thesis. It will be described in more detail in Section 2.6.1.

An interesting addition to the discussion of the previously identified motion mechanisms - transient and sustained (Kulikowski and Tolhurst, 1973; Tolhurst, 1973; Watson and Robson, 1981; Burr and Ross, 1982) - was brought up by Burbeck and Kelly (1981). They found that masking by an orthogonal grating of the same spatial and temporal frequency as the test grating produced substantial contrast threshold elevation when both the test and the mask were of high velocities, but not in any other frequency region. They interpreted this to mean that there exists a mechanism for high-velocity motion. Kelly and Burbeck (1987) measured the orientation selectivity of this high-velocity mechanism using a pattern-adaptation paradigm. The adapting stimulus was a vertical grating, drifting horizontally at the appropriate velocity, while the test stimulus was a horizontal, counterphase-flickering grating. The high-velocity mechanism actually behaved isotropically, i.e. the threshold-elevation ratio was essentially independent of the orientation of the test grating when the spatial frequency was sufficiently low and the temporal frequency was sufficiently high. This isotropic mechanism is broadband in its temporal frequency response, peaking at a high flicker frequency, and is lowpass in its spatial response, cutting of at a few cycles per degree. Kelly and Burbeck concluded that the isotropic mechanism appears to be the same as the so-called transient mechanism. Kulikowski (1991), on the other hand, regards this mechanism as a third, fast motion mechanism, different from the transient mechanism. However, he softens his comment by adding that among these mechanisms, probably only those covering extreme ranges - fast and slow - are subserved by different neuronal mechanisms.

To summarize very briefly the psychophysical observations and their interpretations presented in this chapter, it seems that the visible spatial frequency range can be subdivided into several mechanisms whereas it is possible to identify only very few mechanisms subserving the perception of movement and temporal variations.

2.4. Blurring of images in human vision

2.4.1. Definitions of blur

The images we perceive are not exact pictures of the objects in the physical world. Many factors degrade them: object motion, the optics of the eye, spatially discrete sampling of light distribution by the photoreceptors, processing of these samples by ganglion and other retinal cells, transmission of signals through nerve fibres with restricted dynamic range, and others. Some forms of degradation are useful to the visual system - they can help to avoid aliasing, to improve sensitivity and to diminish the input load of the central processing units (Barlow, 1981). In resolution tasks, however, the degradation always reduces the performance of the visual system.

A mathematical model of a degrading process is as follows: Let $f(x,y)$ denote the object, $g(x,y)$ denote the degraded or blurred image, and $n(x,y)$ denote additive noise. The blurring process can be modelled as

$$g(x,y) = \iint h(x-x',y-y')\, f(x',y')\, dx'\, dy' + n(x,y) \qquad (2.1)$$

The function h is the impulse response or the point spread function of the system. One should note that in the terminology of psychophysics or vision research, additive noise is often excluded from blurring - the effects of noise are either ignored, i.e. they are assumed to be insignificantly small, or they are taken into account separately. Also in this thesis, the term blurring is defined to include only the convolution integral part of Eq. (2.1). If this definition is used, the function h can also be called the blurring function. In the visual system, the blurring functions are combinations of optical and neural blurring. In the previous sections they were called spatial filters or spatial frequency channels. In mathematical terms, blurring is equivalent to spatial filtering - both are produced by convolving the signal with the same function. This meaning of the word blurring is broader than in everyday language: for instance, a difference-of-Gaussians type spatial filter sharpens edges that contain frequencies to which the

filter is tuned, but it blurs or makes the edges containing other frequencies vague.

Equation (2.1) describes the blurring of stationary objects. With moving objects, the velocity of the object and the temporal impulse response of the system must be taken into account. Both of these are functions of time, and theoretically one could write a more general form of Eq. (2.1) by including time as a variable. However, with a uniform, rectilinear motion, it is simpler to transform the effects of motion into the spatial domain and include them in the blurring function.

2.4.2. Blur discrimination studies

Almost all of the psychophysical studies related to blur have examined the effects of blur upon performance. However, blur is not just a degrading factor; it is a measurable quantity of visual percept, and observers can discriminate shadings in illumination such as edge blur with high precision. Hamerly and Dvorak (1981) were the first to measure discrimination thresholds of edge blur. They found that, at high luminance ratios, observers could discriminate a Gaussian-blurred edge from an unblurred edge when the edge transition width (defined as the spatial width over which the luminance goes from 10% to 90% of its range) was of the order of 25 sec of arc. However, when both images were blurred, they could be discriminated when they differed by as little as 5-10 sec of arc. One problem with the interpretation of these results is that the stimulus configuration of Hamerly and Dvorak was such that subjects could use vernier alignment cues, and this may have affected the results.

Watt and Morgan (1983b) measured the thresholds for discriminating the difference in the blur of two edges with various types and extents of blur. They used stimulus settings where subjects could not use cues other than blur extent. Subjects were presented with two edges, and their task on each trial was to report which edge appeared to be more blurred. The sensitivity to extra blur was measured for several different reference blurs. Thresholds for difference in blur extent for two observers, as a function of reference blur for a Gaussian blurring function are shown in Fig. 2.8. The results for rectangular and one half-cycle cosinusoidal blurring functions were similar. The results confirmed the basic findings of Hamerly and Dvorak (1981). Blur comparison is most

precise at some non-zero reference blur for each blurring function - a certain amount of blur is essential in attaining optimal accuracy in this task. Thresholds depend on the reference blur according to a U-shaped function: thresholds decrease as the reference blur is increased from zero to an optimum level, beyond which thresholds rise rapidly. Watt and Morgan found that the rising portion corresponds to a power law with an exponent of 1.5, i.e. thresholds rise faster than they would be expected to from a simple Weber's Law relation. For this thesis, this work by Watt and Morgan (1983b) is fundamental - the main experiment of the thesis is a motion domain extension of their blur discrimination experiment.

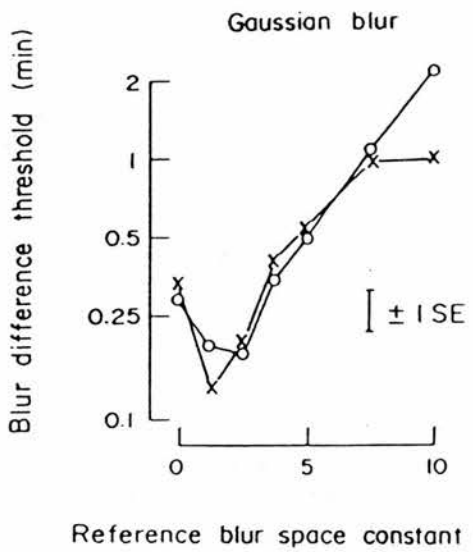


Fig. 2.8. Blur discrimination thresholds for a Gaussian blurring function as a function of reference blur extent. The space constant employed to express blur extent is the standard deviation for the Gaussian. Results are shown for two observers. From Watt and Morgan (1983b).

2.4.3. The effects of blur upon performance

Blur is one of the critical limiting factors in spatial vision. A large number of experiments have studied the effects of blur upon performance in different psychophysical tasks. Among others, these tasks include vernier acuity (Stigmar, 1971; Watt, Morgan & Ward, 1983; Williams, Enoch & Essock, 1984; Toet, van Eekhout, Simons &

Koenderink, 1987), bisection acuity (Toet, van Eekhout, Simons & Koenderink, 1987), spatial separation discrimination (Watt & Morgan, 1983a; Morgan & Ward, 1985; Toet & Koenderink, 1987), stereo-acuity (Stigmar, 1971; Westheimer & McKee, 1980), and displacement detection (Westheimer, 1979; McKee & Nakayama, 1984; Mather, 1987). Stimulus blur affects performance on each of these tasks differently.

For simplicity of the mathematical and psychophysical analysis of the resulting images, most investigators have used ground-glass or Gaussian blur, but even with that the effects of blur are not simple. Gaussian blur will decrease target contrast, and if the performance in the task is sensitive to contrast then the change in contrast alone will degrade performance. The effect of stimulus blur depends on target configuration: in their careful study Williams et al. (1984) found that when the target consists of abutting lines or two points separated by a small gap, i.e. when the stimulus features are placed close together (less than 4 to 8 arc min apart), even a moderate amount of Gaussian blur has a large effect on vernier alignment thresholds. Conversely, they also found that if the gap between the features is large, blur has little effect on thresholds. Stigmar (1971) had noted similar effects of Gaussian blur on stereo- and vernier acuity.

Levi and Klein (1990) performed an elaborate study of the effect of blur with the object of estimating the limiting equivalent intrinsic blur of the visual system. They examined the effect of Gaussian blur on two-line resolution. This task essentially measures the smallest separation between two lines that can be reliably distinguished from no separation. Levi and Klein found that the results at each eccentricity can be summarized by the following equation:

$$Th = k\sqrt{\sigma^2 + B_i^2} \qquad (2.2)$$

where σ is the stimulus blur, specified as the standard deviation of the Gaussian, B_i is the equivalent intrinsic blur, Th is the resolution threshold, and k is a multiplicative constant. This equation is based on a formulation originally proposed by Watt and Morgan (1984). The assumption is that the visual system has an additive internal or intrinsic error that acts like blur (B_i). Since both optical and neural factors contribute this error, Levi

and Klein selected the term equivalent intrinsic blur. The variance of the stimulus blur is expected to add to the intrinsic blur since these two are uncorrelated. The same also follows if we assume that the stimulus blur corresponds to the image, and the intrinsic blur to the blurring function in Eq. (2.1) - variances are additive in convolution. Levi and Klein found that in the fovea, two-line resolution thresholds and the equivalent intrinsic blur are about 0.5 arc min (k was found to be close to unity). They also found that the equivalent intrinsic blur is independent of contrast, and suggested that it reflects the optical quality of the retinal image and discrete sampling by the cones.

Levi and Klein (1990) also examined the effect of blur on line detection and spatial interval discrimination. When the stimulus blur is less than the equivalent intrinsic blur, then the contrast threshold for line detection is approximately inversely proportional to the stimulus blur, indicating complete spatial summation (i.e. Ricco's Law). When the stimulus blur is greater than the equivalent intrinsic blur, then the detection threshold is approximately a fixed contrast. Thus, the equivalent intrinsic blur seems to play a dual role in determining both the resolution threshold and the detection threshold. It corresponds to the diameter of complete spatial summation (Ricco's diameter), and it also corresponds to the resolution threshold for thin lines. Spatial interval discrimination is independent of stimulus blur when the separation is greater than three to four times the standard deviation of stimulus blur, and the thresholds are proportional to the separation of the lines (i.e. Weber's Law). Thresholds begin to rise as a function of stimulus blur when it exceeds about one third to one half the distance between the blurred lines. Thus, the visual system's tolerance to stimulus blur increases with line separation.

There are many studies that have examined the effects of target motion upon performance on different psychophysical tasks. However, these studies are not studies of the effects of motion blur since they have not estimated its amount. Some of those studied will be reviewed in Section 2.5.4.

2.5. Visual motion processing

2.5.1. Functional aspects

The measurement of visual motion is one of the most fundamental abilities of biological vision systems. It has a large number of different roles to play in vision. Most people agree that the leading role is to provide information about moving objects. However, visual motion is also one of the primary sources of information for the moving organism to know about its own motion in relation to its environment, i.e. visual motion serves as a proprioceptive sense (Nakayama, 1985). Motion signals also provide inputs to centres controlling eye movements, allowing moving objects to be tracked. Visual motion perception has a role in the perception of depth and in the segregation of objects. The existence of these diverse functions strongly suggests that several motion systems exist simultaneously. However, it is likely that different functional applications share at least some of their subsystems.

The purpose of this work is to study the possible motion blur prevention system of human vision. This system can be expected to be an integral part of the process that provides information about moving objects. Most of the models for visual motion processing in the literature have also concentrated on modelling how visual system extracts information about moving objects. The following discussion is an attempt to emphasize some of the main functions of the sensing of real motion of moving objects. This separation is of course at least partly artificial, and there are bound to be some tasks shared by different functions. It is not even necessary to use the motion system to solve some of these tasks: for instance, moving objects may be localized using the same mechanism that is used for stationary objects. One purpose of this discussion is to introduce some concepts for use in the next section in the comparison of models for motion perception.

Detection of a moving object

In nature, detection of a moving object is one of the most essential functions of visual motion processing. A sudden movement in the visual scene may be a sign of a dangerous predator or a desirable pray. Our attention is immediately and automatically

directed to the moving object in an otherwise stationary visual field. The same effect on our attention can also be produced by a flash of light, indicating that any change in the scene and not just the movement is the effective factor. However, in natural scenes almost all rapid changes are produced by motion. On the other hand, the background does not even have to be stationary for an object to be almost immediately detected. If the object moves with different velocity than the background, it tends to "pop out" and to attract one's attention. Since all the scene is changing, this effect cannot be based just on the detection of change in one part of the visual field. The mechanism that produces this object-background segregation may be different from the one that detects sudden changes in the visual field.

Localization of a moving object

The most obvious way for the visual system to localize moving objects would be to use the same internal representation that it uses for stationary visual fields. The simplest form for this kind of representation is a retinotopic map. The image and the positional accuracy of the moving object in a retinotopic map are degraded by motion blur, the amount of which depends on both the velocity of the object and the integration time of the mechanism producing the representation. The integration time for stationary objects in human vision is rather long, about 120 ms in daylight (Barlow, 1958). If our visual system uses just this mechanism for localizing moving objects, our perceptual uncertainty in the instantaneous spatial position of a moving object is about one eighth of the distance that the object travels in one second. This value seems to be large, but do we really need greater accuracy? It is more important for us to be able to predict accurately the object's position at a certain moment in the future than it is to know its exact position at the moment of observation. In this prediction the visual system can use information about the past positions of the object along its trajectory. But then again, is it the prediction based on past locations that gives us the location of the object at any moment - and does the integration time correspond to simple retinotopic summation or the time period from which the locations used for the prediction are taken?

Resolving the form of a moving object

If our visual system could take snapshots of moving scenes, we would not need any special motion mechanisms for accurate localization of moving objects nor for resolving the form of their two-dimensional projections. However, temporal summation for moving objects is as strong and as extended as for stationary objects, the total summation time being over about 100 ms (Burr, 1981). Thus, accurate perception of form in motion may require such a visual integration scheme that can summate positional information without smear.

In vision research, the general meaning of the interpretation of structure from motion is the recovery of three-dimensional shape using motion information alone. This capacity of the human visual system has been demonstrated in the classical studies of Wallach and O'Connell (1953). Using two-dimensional shadow projections of an unfamiliar object, they showed that the human visual system can derive a correct description of the hidden object's three-dimensional structure and motion in space, even when each static view is unrecognizable. Other well-known experiments demonstrating this ability were performed by Johansson (1975) with so-called biological motion displays. He showed that a brief observation of patterns of moving lights generated by human figures moving in the dark can lead to a perception of three-dimensional motion and structure of the figures.

One of the computational problems in the interpretation of structure from motion is the ambiguity that infinitely many combinations of three-dimensional structure and motion could give rise to any two-dimensional projection. An additional constraint is needed to rule out all but the most plausible three-dimensional interpretation. Perceptual studies suggest that the human visual system has a tendency to choose a rigid interpretation of moving elements (Wallach & O'Connell, 1953; Gibson & Gibson, 1957; Johansson, 1975, among others). Many computational studies have adopted this rigidity assumption. This constraint assumes that if it is possible to interpret the changing two-dimensional image as the projection of a rigid three-dimensional object in motion, then such interpretation should be chosen - for a review of these algorithms and their more flexible descendants, see Hildreth and Koch (1987).

One remarkable ability of the visual system is to synthesize forms by interpolation. A simple demonstration of this capacity can be produced by sliding an ordinary hair comb on or slightly above this text. The text is clearly visible, though the different parts of the letters are never simultaneously in view.

Deriving the velocity of a moving object

Velocity is one of the most important attributes describing a moving object. It is essential for any hunter - be it a lion or a duck shooter - to be able to anticipate the position of its prey at the moment of hit. This anticipation cannot be accurate without knowledge of the velocity of the prey. When driving a car or walking in the street, we estimate velocities to avoid collisions with other cars and pedestrians. Visual information of the motion of other objects has a fundamental role in the guidance of one's own actions.

An object is in motion relative to another when its position, measured relative to the second body, is changing with time. To describe motion, the observer must define a frame of reference relative to which the motion is analyzed. Since the two-dimensional projection of the outside physical world is mapped onto the retina, the retina could be taken as a candidate for the frame of reference of low level motion perception. In some species, like in the rabbit (Barlow & Levick, 1965), elementary motion measurements are indeed performed at the retinal level. In humans, motion measurements are performed in the visual cortex, and there are principally two different ways of obtaining a similar sensation of the velocity of a moving object. When the observer's eye remains still, the image of a moving object sweeps across the retina. When the observer's eye tracks the object and follows it, the image of the object lies stationary on the retina and the surroundings move. However, in both cases the object is seen to move and not the surroundings. Thus, the physically stationary surroundings of a moving object serve as the frame of reference in human visual motion perception. This question is generally not fully addressed in models of motion perception.

In physics, the instantaneous velocity is defined as the time derivative of the displacement. Velocity is a vector quantity - both the direction and the rate of displacement, i.e. the speed are required for its complete determination. Human motion perception is

velocity perception - when we see motion, we perceive, with some accuracy, both the direction and the speed of the moving object. Any model of motion processing should include an explicit assignment of speed and direction, either in separation or combined in one code of velocity. We are also capable of seeing different motions in different parts of an image. Therefore, any model of our motion system must be able to assign different velocities to different local regions of the visual field.

2.5.2. Models of human visual motion processing

The motion of objects in an image is not given to the visual system directly, but must be inferred from a changing two-dimensional array of light intensity values recorded at the retina. A number of studies have been concerned with developing models for motion detection and measurement by biological visual systems (Reichardt, 1957; Barlow & Levick, 1965; Marr & Ullman, 1981; van Santen & Sperling, 1984; Adelson & Bergen, 1985; Watson & Ahumada, 1985; Harris, 1986). The schemes proposed for motion detection have been divided into two main categories, the so-called correlation- and gradient-type models. The current neurophysiological and psychophysical evidence does not decisively favour either of these main schemes.

2.5.2.1. Correlation models

In order to signal motion in a directionally selective way, a motion detector has to be asymmetrical and needs at least two nonlinearly interacting inputs (Poggio & Reichardt, 1973). In correlation models, the nonlinearity is produced by the multiplication operation. In these models, an estimate of local motion is obtained by evaluating a kind of spatiotemporal cross-correlation of the appropriately filtered signals originating from two points in the retinal image (Borst & Egelhaaf, 1989).

The best known and the first explicit model for motion detection was proposed by Reichardt (1957; 1961). This model was formulated to explain optomotor responses in insects. The basic operations of Reichardt's model - and correlation-type models in general - are summarized in Fig. 2.9. In the simplest form, the retinal light intensity distribution is sampled by two receptors with point-like

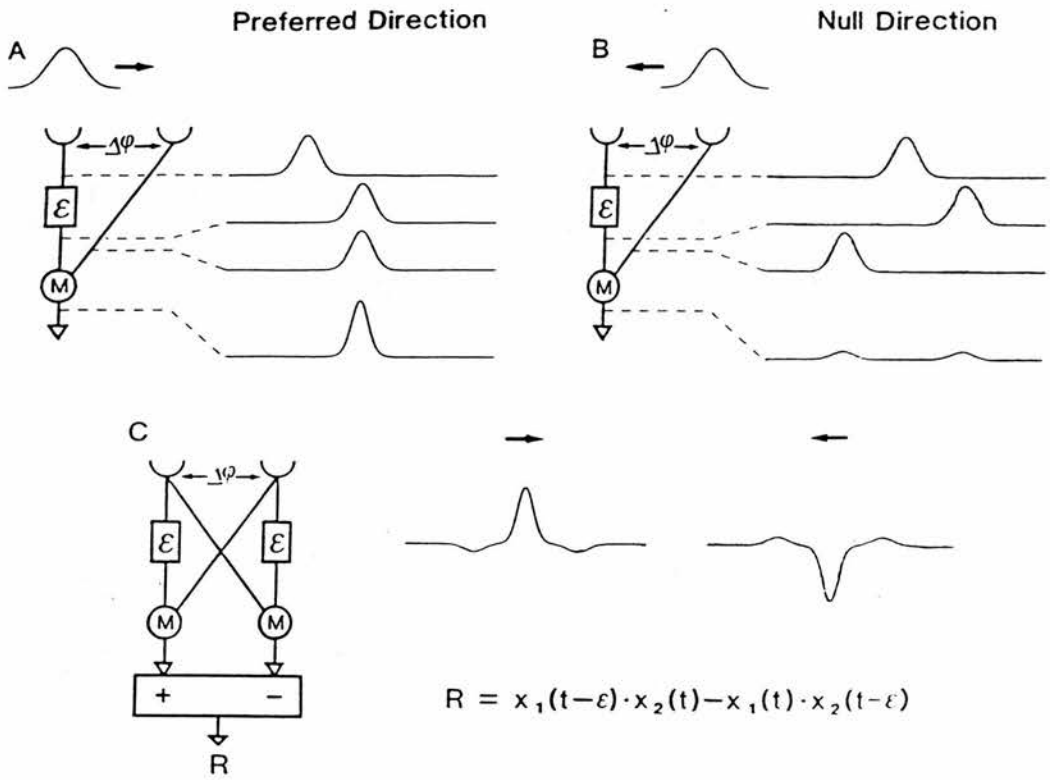


Fig. 2.9. Operating principles of a correlation-type visual motion detector. This detector is composed of two mirror-symmetrical subunits (C). Its input is given by the light intensities measured at two points. In each subunit, the input signals are multiplied (M) by each other after one of them has been delayed by a time interval e. The final output of the detector is given by the difference between the subunit outputs. To aid an understanding of the detector's operation, the output of one subunit to motion in opposite directions is considered first. (A) When the stimulus moves in the detector's preferred direction, the delay in the left arm of the subunit may compensate the temporal separation of the signals in the input channels. In this way the signals may coincide at the multiplication stage, resulting in a large output signal. (B) If the stimulus moves in the 'null direction', the delay further increases the separation of the signals, resulting in two small response peaks. (C) By subtracting the output signals of two mirror-symmetrical subunits, the response components due to correlated input signals from background luminance are eliminated. With a perfectly balanced subtraction stage, the responses to motion in opposite directions have the same amplitude and time course but opposite signs. From Borst and Egelhaaf (1989).

receptive fields. The receptors are spatially separated by the sampling base $\Delta\varphi$ of the detector. The signals from the receptors are multiplied (M) by each other after one of them has been delayed by a time interval ε . If the delay is appropriate, the signal in the input channel which is activated first by a moving stimulus coincides with the other channel's signal at the multiplication stage. This results in a large response amplitude (Fig. 2.9.A). On the other hand, if the stimulus moves in the opposite direction, the delay further increases the separation of the signals. Both of the signals are multiplied only by the background luminance level, resulting in small responses (Fig. 2.9.B) if the background level is low. In reality, the response components induced by correlated input signals from stationary background objects may be substantial. To eliminate these, a correlation-type motion detector consists of two subunits that are mirror images of each other. The final output is the difference between the subunit outputs (Fig. 2.9.C). The responses to motion in opposite directions have the same amplitude and time course but opposite polarities.

Reichardt's model predicted a number of features that were in accordance with the data obtained from insects. One of the most interesting is motion inversion, resulting from spatial aliasing: If the wavelength of a periodic stimulus pattern is less than twice the sampling base or the separation between the input channels, the insect will perceive motion in the direction opposite to the true direction of motion (Reichardt, 1969).

Barlow and Levick (1965) proposed an alternative method to account for the behaviour of directionally selective units in the rabbit's retina. This model follows the same general lines as Reichardt's model - the multiplication operation is just replaced by an AND-operation and a NOT-operator is added to the delayed channel. The change from analogue operations to logical operations was inspired by evidence for inhibitory interactions within the directionally selective mechanism. Barlow and Levick demonstrated that it was inhibition in the nonpreferred direction and not excitation in the preferred direction that played the major role in directional selectivity; inhibition vetoed the responses to movement in the nonpreferred direction.

It is generally accepted that early processing of visual information in the human visual system contains channels responding to different spatial frequencies of the image. Along these lines,

van Santen and Sperling (1984) proposed a modified and elaborated version of Reichardt model in which the point-input assumption is generalized to the input of the entire stimulus through a linear spatial filter or spatial-frequency-selective receptive field. The major advantage of the front-end spatial filter is that it eliminates the spatial aliasing of the original Reichardt model - humans do not show spatial aliasing in normal viewing. Van Santen and Sperling tested their model with a number of psychophysical experiments. Specifically, they found that if the contrast of neighbouring vertically oriented bars moving in a horizontal direction is varied, performance depends on the product of adjacent bar amplitudes, offering strong support for the multiplication principle.

Adelson and Bergen (1985) proposed not a particular model but a class of models that arise from a simple spatiotemporal conceptualization of motion. A moving two-dimensional pattern can be described as occupying a three-dimensional space, where x and y are the two spatial dimensions and t is the temporal dimension. A motion sequence may be presented as a single pattern in this space. A velocity corresponds to a three-dimensional orientation in x-y-t space. Motion information can be extracted by a system that responds to the oriented spatiotemporal energy. Adelson and Bergen showed how such a system can be built by combining space-time separable impulse response functions. Separability has been observed in some neurophysiological studies (Movshon, Thompson & Tolhurst, 1978) but there are others that consider it at least an oversimplification (Fleet, Hallett & Jepson, 1985; Burr, Ross & Morrone, 1986a). However, the separable spatiotemporal impulse response is the simplest way to combine the spatial and temporal impulse responses. Fig. 2.10 shows an example of this kind of construction. Across the top there is a profile of a difference-of-Gaussians-type spatial impulse response. A typically biphasic temporal impulse response is shown running down in the vertical direction. The spatiotemporal impulse response is the product of the spatial and the temporal impulse response. In this case, the result contains six lobes, alternately positive and negative. In mathematical terms, the output of the unit in Fig. 2.10 is the temporal convolution of the impulse response with the spatiotemporal input pattern. A row of similar units carries out a convolution in both space and time. The resulting mechanism or channel performs a

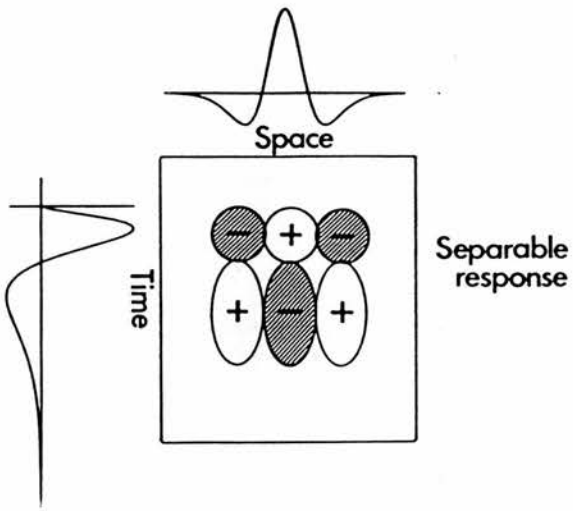


Fig. 2.10. A spatiotemporally separable impulse response (shown schematically in the centre) is the product of the spatial and temporal impulse responses (shown along the margins). It is a weighting function that sums inputs at various positions and times to determine the present output. From Adelson and Bergen (1985).

filtering operation that selectively passes some of the spatiotemporal energy of the stimulus.

The unit in Fig. 2.10 will respond strongly when there is motion within its receptive field but will not respond to non-moving targets. Thus, it would be a good candidate for a motion detection, except that it is insensitive to direction of motion. This is a consequence of that the unit has no dominant spatiotemporal orientation. On the other hand, a unit that does have spatiotemporal orientation will be selective for direction of motion. Oriented filters, however, do have their own problem: they are phase-sensitive, that is their response to a moving pattern depends on how the pattern lines up with their receptive field at each moment. For instance, a moving sine-wave grating produces a response that itself oscillates sinusoidally over time.

A phase-independent motion detector can be built by combining two oriented linear filters that are 90 deg out of phase, i.e. that form a quadrature pair. This is shown schematically in Fig. 2.11. The filters are superimposed in both space and time, and so the squares in the figure represent the same spatiotemporal region. To extract

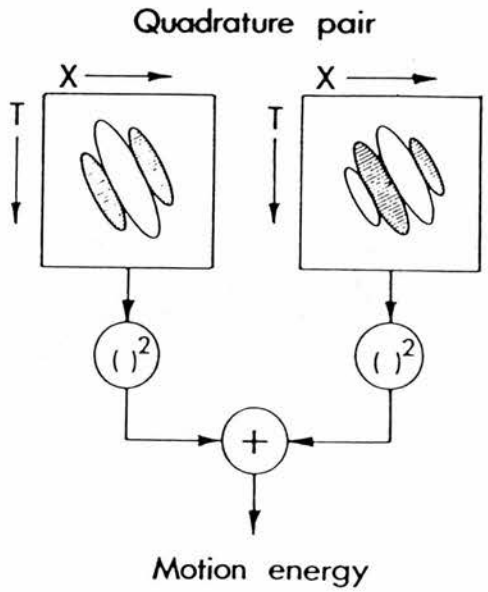


Fig. 2.11. A schematic diagram of a spatiotemporal energy detector. A quadrature pair is formed by superimposing two Gabor functions that are 90 deg out of phase (i.e. sine and cosine functions that are weighted by the same Gaussian window). The responses of the filters are squared and summed to extract a phase-independent measure of local motion energy. From Adelson and Bergen(1985).

a measure of local motion energy, the outputs of the filters are squared and summed. Leftward and rightward energy detectors may be combined to produce an opponent detector. The system gives a steady response to a steady motion, and the sign of the response depends on the direction of the motion and not on the polarity of the stimulus. A spatiotemporal energy detector is a local motion detector in many respects: its response is localized in both space, time, and spatial frequency. Adelson and Bergen (1985) used their model to explain many basic phenomena and various illusions in motion perception.

Watson and Ahumada (1983) proposed a motion detector that replaced the multiplier in the Reichardt model with an adder, resulting in a linear detector. In this model, directional selectivity is accomplished by arranging a pair of spatially tuned receptive fields in spatial and temporal quadrature for all spatial and temporal frequencies. This model is closely related to the elaborated Reichardt model of van Santen and Sperling (1984). In fact, van Santen and Sperling (1985) showed that with suitable chosen filters, their model is fully equivalent to the model of Watson and Ahumada. They also showed that for every choice of

filters, the elaborated Reichardt detector is fully equivalent to the spatiotemporal energy model of Adelson and Bergen (1985).

Most of the models of motion perception have given explicit theories of direction detection but they do not have any feature for metrical velocity coding. This is partly a result of a general property of correlation-type motion detectors: their output depends on the structure of the stimulus pattern, such as the spatial frequency content and contrast (Reichardt, 1961; Borst & Egelhaaf, 1989). While the elaborated Reichardt detector (1984) can detect motion and discriminate the direction of motion, it cannot reliably estimate velocity. For the spatiotemporal energy model, Adelson and Bergen (1985) proposed a system where the activity of the sensors responding to opposite directions can be used to compute a measure of velocity that is invariant with contrast. They did not, however, make any attempt to make the computations of metrical velocity values explicit. Watson and Ahumada (1985) developed a two-stage model that gives a more comprehensive treatment of velocity. The first stage of this model is a set of spatial-frequency-tuned, direction-selective linear sensors. Watson and Ahumada showed that the temporal frequency of the response of each sensor encodes the component of the image velocity in the sensor direction. At the second stage, these components are resolved to measure the velocity of image motion at each of a number of spatial locations and spatial frequencies. Effectively, this model estimates velocity by measuring temporal frequency and dividing it by the preferred spatial frequency of the channel. Thus, no special velocity channels are used in this model.

In all the models described above, motion measurements are based on local changes of light intensity values, and thus these models can be called intensity-based schemes. On the other hand, more symbolic features such as zero-crossings, edges, corners, blobs, or regions could be used as motion primitives. The models that measure motion by matching these features over time are called token-matching schemes (Ullman, 1983). The apparent-motion phenomena demonstrate the ability of the visual system to establish visual motion by matching tokens over considerable distances in space and time. If we assume that the input to the visual system is given as two successively presented frames, the visual system is faced with the problem of locating a counterpart for each element in the first frame in the second. This so-called correspondence problem is illustrated in Fig. 2.12. The filled squares in the

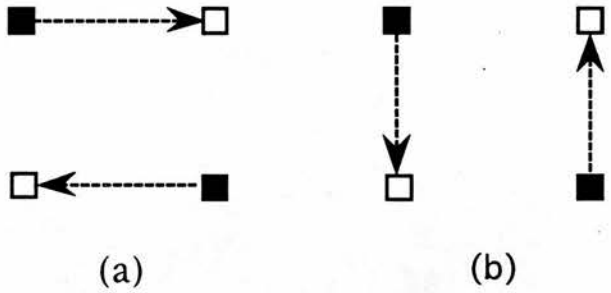


Fig. 2.12. A simple correspondence problem. Two frames are presented in brief succession. The filled squares represent the first frame and the open squares the second. There are two possible one-to-one matches between the elements of the two frames, leading to two patterns of perceived motion: clockwise (a) and counter-clockwise (b).

figure represent the first frame and the open squares the second. There are two possible one-to-one matches between the elements of the two frames, leading to two patterns of perceived motion: clockwise and counter-clockwise. In this example, the match is only two-way ambiguous. In general, each frame may contain many elements arranged in complex figures, and a correspondence has then to be established between the figures. Complex tokens can simplify the correspondence process by making the match less ambiguous, but more computation is required to extract these features from the image. Ullman (1979) proposed a token-matching scheme for the measurement of apparent motion, which he called the Minimal Mapping Theory. In this method, features in a given frame are matched to features in a second frame in such a manner that the sum of the distances travelled is minimal. By distances, Ullman meant not only Euclidean distances, but also more abstract notions of distance, such as the difference in orientation or in brightness of features. Ullman's model assumes independence of the matching elements. Recent studies have revealed situations where this assumption appears not to hold. Ramachandran and Anstis (1985) performed experiments with a display in which a local pattern of dots whose motion was two-way ambiguous was repeated in a large array. They found that observers always perceived the array of patterns as moving in the same direction. Thus, the correspondence established within one subpattern of the display influenced the correspondence of dots in neighbouring subpatterns.

Solving the correspondence problem is generally assumed to be one of the fundamental issues in visual motion perception - for a review see Hildreth and Koch (1987). Failure to solve this problem will lead to false motions between non-corresponding objects, or to the breakdown of motion percepts (Anstis, 1983).

2.5.2.2 Gradient models

In gradient schemes, the local motion measurements are derived via a comparison between the spatial and temporal gradients of image intensity. In mathematical terms and in the case of one-dimensional movement, the local velocity dx/dt is obtained by dividing the temporal gradient dI/dt by the spatial gradient dI/dx of the pattern (x and t refer to the spatial variable and time, and I denotes the light intensity. In two dimensions, only the component of the velocity in the direction of the brightness gradient can be measured. This limitation, termed the aperture problem, is illustrated in Fig. 2.13. If the motion of the edge is to be detected by sensors which examine an area A that is small compared to the overall extent of the edge, the only motion that can be extracted is the component c perpendicular to the local orientation of the edge. No distinction can be made between motion in directions b, c, and d. A subsequent processing stage, which integrates the local measurements, for instance along the contours of the object, is required to determine the motion completely.

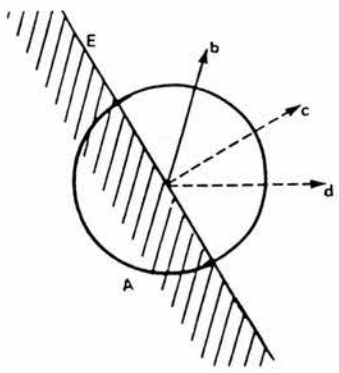


Fig. 2.13. The aperture problem. Looking through the local aperture A, it is impossible to distinguish between motion in direction b, c and d. From Ullman and Hildreth (1983).

Marr and Ullman (1981) proposed a quantized version of the gradient scheme, which was motivated by computational studies of early visual processing and by neurophysiological studies of directionally selective simple cells in primate visual cortex. According to this model, the visual system detects the presence of an edge in the retinal image and concurrently registers the direction of change in the luminance at the location of the edge. Marr and Hildreth (1980) have shown that edges can be found by convolving the original image with a Laplacian of a Gaussian, whose shape may be approximated by the difference of two Gaussians, i.e. by the centre-surround receptive field. The elements of the convolution output, which correspond to the locations of sharp intensity changes in the image, are the zero-crossings. If the values of the convolution are carried out by two kinds of unit, one dealing with positive values (S^+ or on-centre unit) and the other with negative values (S^- or off-centre unit), S^+ units will be active on one side of the zero crossing, and S^- units on the other side. If the outputs of the two sides are fed to a logical AND gate, the gate will give a non-zero output only if there is a zero-crossing with the correct edge polarity between the sides. A row of such units will detect an oriented segment of zero-crossings. A second type of detector, termed a T unit, samples the temporal derivative of the convolution in the same patch of the image as the edge-detecting unit. Also the temporal unit is signed to indicate whether it gives an output to an increase (T^+ unit) or to a decrease (T^- unit) in luminance. A directionally selective unit can be constructed by passing the output of either a T^+ or T^- unit to a logical AND gate together with the outputs of a S^+ and S^- unit. This yields a set of STS detectors signalling the leftward or rightward motion of bright-dark or dark-bright edges.

Marr and Ullman tentatively identified the edge detecting S^+ and S^- units with sustained X-like on- and off-centre retinal ganglion cells and the T+ and T^- units with transient Y-like on- and off-centre cells. They presented a number of neurophysiological predictions derived from their theory that are consistent with experimental findings. However, some aspects of the proposed implementation are contradicted by neurophysiological accounts of neural mechanisms. For example, Hochstein and Shapley (1976) found that Y-like cat retinal ganglion cells are fed by rectifying subunits. Thus, they are unlikely to transmit the sign of the temporal derivative. In psychophysics, there is one study that

provides strong evidence in favour of the Marr-Ullman model. Moulden and Begg (1986) demonstrated the existence of adaptation which is specific not merely to the direction of movement of an edge but also to its contrast polarity. In addition, they showed that adaptation to a nonmoving, spatially homogeneous field whose luminance is modulated according to a temporal sawtooth waveform produces changes in sensitivity to the movement of an edge that can be predicted on the basis of the Marr-Ullman model. The visual system seems to contain adaptable units that are specifically tuned to signal a particular change in luminance.

In contrast to other gradient schemes, the Marr-Ullman model does not provide an estimate of the local velocity, but only its sign or the direction of motion. Harris (1986) has proposed an extension of the Marr-Ullman model for direction coding that is more faithful to the original gradient scheme principle. In his model, velocity is computed within each spatial channel by taking the spatial and temporal derivatives of the Laplacian of a Gaussian filtered image and dividing the latter derivative by the former.

2.5.3. Forms and phenomena of visual motion

Bound by the laws of dynamics, natural objects of the physical world tend to move smoothly and continuously. With the exception of objects occluding each other, their visual motion can be described with the same attributes. However, our perception of motion is not restricted to continuously moving objects - the visual system can perceive motion when nothing actually moves.

Apparent motion

An illusion of continuous motion can be produced by displaying still pictures of a moving object in sequence. This illusion is called apparent motion to distinguish it from "real" motion, which we perceive when an object moves continuously across our visual field. This illusion is more common than we normally think - everyday examples of it include cinema and television. Apparent motion was found by psychologists experimenting with stroboscopes in the 19th century. When a stroboscope is flashed sequentially in darkness, observers see static images of a moving object. However, if the time intervals between the flashes are short enough, a

compelling impression of continuous motion is perceived. In psychophysical terms, stroboscopic or sampled motion is indistinguishable from smooth motion if the physical sampling rate is adequate to include all those frequencies to which the visual system can respond (Morgan, 1979; Morgan, 1980a; Morgan, 1980b; Fahle & Poggio, 1981; Burr, Ross & Morrone, 1986b; Watson, Ahumada & Farrell, 1986). Stroboscopically illuminated targets are seen to occupy positions between those where they are actually exposed. This spatiotemporal interpolation effect has important relations to motion blur, and it is discussed in more detail in Section 2.5.5.

There are plenty of experimental variants of apparent motion in the literature. One of the most quoted observations is from the study of classical apparent motion or "phi" by Wertheimer (1912). He found that an "object-less" movement is perceived between two presentations of an object if the spatial and temporal displacement is suboptimal. This is different from stroposcopic motion in that the object is not seen in positions between the presentation points. Wertheimer did not regard this movement as a perception, instead he termed it a "phenomenon" - phi is the abbreviation of that term. Wertheimer assumed that phi originates from the electrochemical connections between the cortical projection points of the stimulus. This interpretation suggested that a sequence of images might mimic the encoding of real motion by the brain, and that the appearance of movement was something added to the sequence by the observer. Though there have been several explanations of apparent motion based on the supposition that motion is derived from stationary frames by cognitive (Johansson, 1975) or computational processes (Ullman, 1979), this view is now generally assumed to be an over-simplification (Nakayama, 1985; Burr, Ross & Morrone, 1986b).

Short-range and long-range motion

Many studies suggest that there are two types of process that underlie the appearance of motion, one with a short and the other with a long range. The short-range process is thought to occur at a relatively low level of visual processing, and the long-range range process at a higher, cognitive or interpretative level. With respect to the measurement of visual motion, the short-range process is very likely to be an intensity-based scheme, while the long-range process appears to be a token-matching scheme.

The distinction between short-range and long-range motion processes was first proposed by Braddick (1974). He displayed a pair of random dot patterns in rapid alternation. The dots in a rectangular area of one of the patterns were identical to those in the corresponding area of the other pattern but displaced slightly. The rectangular area itself remained stationary. This type of display yields an impression of a rectangle oscillating back and forth against the background if the patterns are displayed at suitable intervals. Braddick measured the maximum displacement (D_{max}) at which directional motion could be reliably perceived, and got an estimate of approximately 15 min of arc. He also found that the perception of the oscillating square deteriorated as the dark interval separating successive frames approached 100 ms. Braddick concluded that there exists two types of motion mechanisms: a short-range process that operates over brief durations and distances up to 15 min of arc and a long-range process operating over long distances and long durations. Subsequently it has been demonstrated that D_{max}, the limit of the short range process, is not an absolute but a relative limit. Baker and Braddick (1985) redefined it as a measure that remains constant at any given retinal eccentricity. However, there is direct and indirect evidence that D_{max} does not remain invariant even within a given eccentricity, but varies as a function of the spatial frequency content of the stimuli (Lappin & Bell, 1976; Chang & Julesz, 1983; Cavanagh, Boeglin & Favreau, 1985; Burr, Ross & Morrone, 1986b; Bischof & Di Lollo, 1990; Boulton & Baker, 1991; Morgan, 1992, among others). Thus, D_{max} can not be regarded as a sound base for the classification of short-range and long-range motion processes. On the other hand, the findings of the temporal range still support the distinction. Baker and Braddick (1985) found that short-range motion detection requires portions of each exposure to occur at about 40 ms separation in time. In contrast, long-range apparent motion can be seen in alternating figures for time separations up to several hundred milliseconds (Kolers, 1972; Burt & Sperling, 1981; Mather, 1988).

First-order and second-order stimuli

Cavanagh and Mather (1989) argue against the distinction between short-range and long-range motion processes from a different point of view. They claim that the observed differences between these phenomena are not evidence for the existence of two different motion

processes but a consequence of the stimuli used in the two paradigms. They propose another dichotomy based on stimulus attributes: first-order and second-order stimuli. Cavanagh and Mather identify colour and luminance as first-order attributes, and properties such as texture, motion and binocular disparity as second-order attributes. They argue that different motion detectors selective for various first- and second-order stimulus attributes share a common mode of operation, which is a dense array of comparator-type motion detectors.

The perception of motion in the absence of correlated luminance information suggests the existence of second-order motion. Among the first to demonstrate this phenomenon were Ramachandran et al. (1973), who reported a clear impression of motion in a two-frame display consisting of two completely uncorrelated random dot patterns. Each frame had a square region of random visual noise on a background that consisted of different random noise, having the same mean luminance but different second order statistics. The position of the square in one frame was displaced slightly relative to its position in the other. When the frames were presented in alternation, an impression of back and forth motion of the square was produced. Recently, Albright (1992) presented neuro-physiological evidence that many motion-sensitive neurones in the middle temporal area (MT) of monkey visual cortex have similar directional tuning for both first- and second-order motion. The second-order stimulus used in this study had the appearance of a rectangle of flickering dots that drifted smoothly across a background of identical but static texture. The spatiotemporal Fourier spectrum of this stimulus is zero outside the spatial-frequency axis, i.e. there is no oriented spatiotemporal energy in the stimulus. Motion in this kind of stimuli cannot be detected by low-level, luminance-based spatiotemporal-energy or other correlation-type or gradient-scheme-type detectors. Chubb and Sperling (1988) have referred to such stimuli as non-Fourier motion, but they have also noted that the essential features of low-level motion detectors can be applied for motion detection in these stimuli by adding appropriate input filters, for example a flicker detector.

Component and pattern motion

Human motion percepts are not always unambiguous. For example, when two drifting gratings with different orientations are superimposed to form a plaid pattern, there are two possible impressions of motion. The observer may either see a rigid coherent cross-hatched pattern moving in a single direction or he may see the two component gratings moving transparently and independently in different directions. Adelson and Movshon (1982) suggested that if coherent pattern motion is perceived, the direction and the speed of the pattern can be accounted for by a geometric combination rule known as the intersection-of-constraints. When two constraint lines are drawn so that they are orthogonal to the component motion vectors, the intersection of the lines gives the resultant pattern motion. The result is not a vector sum or vector average, instead it is a simultaneous solution to the aperture problem for both of the component motions. With asymmetric plaids, deviations from the predictions of the intersection-of-constraints computation have been found (Kooi, Grosof, DeValois & DeValois, 1988; Ferrera & Wilson, 1990). Ferrera and Wilson (1991) suggest that the basic reason for these deviations may be the responses of low-level motion detectors which confound speed with contrast and spatial frequency.

Adelson and Movshon (1982) studied the conditions under which coherent pattern motion is perceived, and found that it depends on the relative contrasts, spatial frequencies and directions of motion of the gratings. In addition, they found that coherence is strongly reduced when a rapidly changing pattern of random-width lines is superimposed on the plaid pattern parallel to the component orientation. This indicates that some orientation-selective process must precede the analysis of coherent motion. They proposed that the synthesis of two-dimensional pattern motion is accomplished in two stages. The first stage consists of one-dimensional motion detectors which signal the motion of the components, and the second stage consists of a two-dimensional motion analyzer which computes the resultant pattern motion. This two-stage model of motion analysis has gained both psychophysical (Welch, 1989; Welch & Bowne, 1990; Derrington & Suero, 1991) and neurophysiological (Movshon, Adelson, Gizzi & Newsome, 1985) evidence and is today widely accepted.

Motion capture

One of the most surprising phenomena of visual motion is the perceptual illusion termed motion capture. If two uncorrelated random dot patterns are alternated, they generate random incoherent motion, like 'snow' in a detuned television set. However, if a moving sine-wave grating is superimposed on this display, all the dots in it appear to move as a uniform sheet in synchrony with the grating - as though they were captured by it (Ramachandran & Inada, 1985; Ramachandran & Cavangh, 1987). Motion capture is one of the illusions that have inspired development of computational theories for the perception of coherent visual motion. Yuille and Grzywacz (1988) proposed a two-stage motion coherence model where local velocities are measured in the first stage and the second, smoothing stage constructs a velocity field over the entire image. The plausibility of the smoothness constraint rests on the principle that velocity varies smoothly across any natural visual field (Horn & Schunk, 1981; Hildreth, 1984). Bülthoff et al. (1989) proposed a parallel algorithm for real-time computation of optical flow. This model predicts and demonstrates motion capture as a result of a local summation or excitation step in the algorithm. Bülthoff et al. generalize this principle by stating that phenomena such as motion capture are to be expected by any algorithm that integrates information about motion over local spatial neighbourhoods.

2.5.4. Effects of target motion on hyperacuity

There is a huge quantity of papers related to visual motion processing in the literature of psychophysics, and it would be an unworkable task to produce a comprehensive review of it. Thus, this and the following section are confined to those areas of visual motion psychophysics that are most closely related to motion-deblurring or motion-blur prevention, namely hyperacuity of moving targets, spatiotemporal interpolation, and temporal summation of moving images.

Under appropriate conditions, human observers can judge the relative position of two visual features with a precision that is far beyond the resolution limit and thus substantially smaller than the size of a foveal cone. Westheimer (1975) called these very precise spatial discriminations 'hyperacuities'. The best-known hyperacuity is vernier acuity, where observers can reliably discern

the direction of an offset of 2-5 arc sec between two abutting lines. Other examples include orientation, curvature, spatial interval or width, bisection, and stereoacuity. Hyperacuity also occurs in foveal motion sensitivity - the smallest movement or displacement (D_{min}) that can be detected has been found to be about 5 arc sec (Nakayama & Tyler, 1981).

Hyperacuities have been used to study the spatial localization capabilities of the visual system, and also for moving targets. Westheimer and McKee (1975) showed that a vertical vernier target could be moved horizontally at velocities up to 4 deg/sec without a reduction in acuity. The vernier threshold was the same, 6 arc sec, whether the target was stationary or moving during a 100 ms exposure. This exposure is too short to permit accurate pursuit eye movements. The insensitivity of vernier acuity to motion is surprising - one would expect from the poor temporal resolution of the visual system that motion blur should degrade the thresholds. Westheimer and McKee (1977) ruled out the 'neural snapshot' explanation, i.e. that the effects of blur would be minimal because the horizontal offset information would be based on a single brief glimpse of the moving target. They showed that information about target configuration was integrated over a distance extending at least 4 min of arc. In their series of studies of hyperacuity and motion, Westheimer and McKee (1978) reported that stereoscopic acuity has a similar insensitivity to target motion as vernier acuity has.

Morgan, Watt and McKee (1983) measured vernier acuity under different conditions of target motion and exposure duration. Their results in the brief exposure condition agreed with Westheimer and McKee (1975): image motion up to about 3 deg/s had little effect upon threshold. In their unrestricted exposure condition the stimulus swept from right to left through a distance of 90 arc min, with an interval of 200 ms between sweeps. In this condition, the threshold for a stationary stimulus was about a third of the threshold for a briefly presented stationary stimulus, and the threshold increased with target velocity until it reached the value of the brief exposure case at about 3.0 deg/s. Pursuit eye movements cannot have caused this effect, since they would predict the reverse effect. Morgan et al. concluded that motion per se decreases vernier acuity relative to its optimum value with a stationary, long-exposure stimulus.

Morgan and Benton (1989) confirmed the finding of Westheimer and McKee (1975) by showing that observers can detect a vernier cue of less than 10 arc sec, and that acuity is little affected by target motions up to 6 deg/s. But, they also found that spatial interval discrimination thresholds for bar separation of 4.5 arc min rise about threefold over the velocity range from zero to 6 deg/s. Morgan and Benton suggested that this difference results from the difference in the spatiotemporal structures of the stimuli. In the two-bar interval case, the bars follow one another rapidly at the same retinal location. The sluggish temporal resolution of the visual system causes their responses to overlap, and their peaks become difficult to discriminate. Morgan and Benton predicted from this that spatial interval acuity would be less affected by motion if the separation between the bars were increased, and their results confirmed this prediction.

In an important study, Welch and McKee (1985) examined the effects of motion on different vernier configurations. The two components of a vernier target were separated and travelled toward each other along trajectories which differed in direction. Preliminary results revealed that in the arrangement where the components collided and their paths crossed, thresholds were very high. Therefore, Welch and McKee chose to use the converging configuration with target trajectories that did not cross. They found that if the components are moving in directions which differ in angle by more than 15 degrees, vernier thresholds rise significantly at speeds greater than 1 deg/s. When the speed goes above 1.5 deg/s, thresholds are never less than 30 arc sec. Welch and McKee tested several conditions which affect stationary vernier acuity, i.e. separation of target components, exposure duration, and orientation, and found that none of these accounts for the loss in acuity. They concluded that directionally-selective motion detectors are responsible for the localization of moving targets.

2.5.5. Spatiotemporal interpolation

The human visual system has a capacity to interpolate almost perfectly the momentary position of a discontinuously moving target, given that spatial and temporal intervals between the stations of actual presentation are small. The precision at which a stimulus can be located at positions between the stations is in the hyperacuity range. Space and time are interchangeable: for

instance, a temporal delay in the presentation of spatially aligned components of a vertical vernier target is perceived as a spatial vernier offset, identical to their simultaneous presentation at spatially non-aligned positions. Morgan (1976) showed that the apparent offset produced by a temporal delay could be cancelled by a spatial offset in the opposite direction. Burr (1979) went on to show that acuity for detecting the illusory offset produced by a temporal delay was very fine, almost as fine as that for detecting real spatial vernier offsets, i.e. the threshold temporal delay for detecting the apparent offset, multiplied by the velocity, was not much higher than the threshold for an explicit spatial offset. It appeared as if the visual system was translating a temporal signal into a spatial signal by using knowledge of the velocity of the target.

Morgan (1979; 1980a; 1980b) examined the temporal limits of spatiotemporal interpolation. He found that the spatial offset equalled the amount the targets would move during the temporal delay, provided that the temporal interval between stations did not exceed about 30 ms. The accuracy of interpolation declined with longer intervals, so that with 50 ms the stations were seen in their actual physical alignment. In explanation of spatiotemporal interpolation, Morgan proposed that it is produced by spatiotemporal filtering, without any special mechanisms for interpolation.

Fahle and Poggio (1981) examined the spatial limits of spatiotemporal interpolation. Their observation was that the distance between the stations of presentation had very different effects on spatial and temporal thresholds: the spatial thresholds were hardly affected by the inter-station distance, while the acuity for temporal offset was impaired almost in proportion to it, i.e. the efficiency of interpolation declined with increasing inter-station distance. Fahle and Poggio proposed a model to account for their findings and for spatiotemporal interpolation in general. The model incorporates 'stasis' channels sensitive to high spatial frequencies and low temporal frequencies, and 'motion' channels sensitive to low spatial and high temporal frequencies. The 'stasis' channels are supposed to detect actual spatial offsets in stationary or slowly-moving targets, while the 'motion' channels would be responsible for temporal interpolation.

Fahle and Poggio (1981) used a signal-analytical approach to explain how interpolation is accomplished in their model. A target moving continuously and with a constant velocity is represented by a

diagonal line in the spatiotemporal Fourier domain, i.e. in a coordinate system of spatial and temporal frequencies. This diagonal line is the spatiotemporal Fourier spectrum of the target. A discontinuous presentation or sampling of the stimulus introduces side lobes on both sides of the line. Interpolation is equivalent to eliminating these side lobes by a filter matched to the diagonal line, that is to the constant velocity of the target. This elimination makes the Fourier representation of the discontinuous stimulus indistinguishable from the representation of the continuous stimulus. It is the easier the larger the distance between the line and the sidelobes, i.e. the denser the spatial and temporal sampling.

Morgan and Watt (1983) carried out a further analysis of the spatial sampling limits of spatiotemporal interpolation, and found that there was a pronounced nonlinearity in the relation between the acuity for temporal offset and the inter-station distance, in contrast to the conclusion of Fahle and Poggio (1981). The distance between the stations had little effect on the temporal threshold until it exceeded about 4.5 min. Morgan and Watt pointed out that this value was strikingly similar to the integration region for moving targets described by Westheimer and McKee (1977). They argued that a model based on separable spatial and temporal filters alone is sufficient to explain the behaviour of spatiotemporal interpolation. The distribution of the output activity in a two-dimensional retinotopic array of such separable filters can be considered as representing the spatiotemporally averaged image of the target. In this image, a temporal offset between the components of a moving target will manifest itself as a spatial offset. If the sampling of the discrete presentation of the target is dense enough, the image will not differ from the one that would be produced by a similar but continuously moving target, and thus a complete temporal interpolation will be achieved. Interpolation will fail if the stations become resolved either spatially or temporally.

Morgan and Watt (1983) tested their ideas with a simulation model of the breakdown of spatiotemporal interpolation. In this model, they used a simple temporal filter described by Roufs and Blommaert (1981) . For a spatial filter, they selected the smallest transient channel , the "T" channel, of Wilson and Bergen (1979). This filter is a difference-of-Gaussians filter, with standard deviations of 5 min and 7.95 min for the excitatory and inhibitory Gaussians, respectively. Their simulation illustrated that interpolation by

this spatial filter was efficient when the sampling period was less than about 5 min, but declined rapidly with larger sampling intervals. Thus, the simulation showed a marked similarity to their psychophysical findings.

The most fundamental difference between the models of spatiotemporal interpolation presented here is that the model of Fahle and Poggio uses motion-tuned detectors or filters which are non-separable in space and time, while the model of Morgan and Watt is based on separable spatial and temporal filters which are not tuned to motion.

2.5.6. Temporal summation of moving images

The visual system does not respond to instantaneous distributions of light imaged on the retina, but summates signals over time. Bloch (1885) showed that the visibility of a flash of light depends only on the light energy of the flash. If the luminance of a dot is kept constant over time, the perceived brightness of the dot depends linearly on the exposure duration. Later, it was shown that there is an upper limit for this complete temporal summation (Graham & Margaria, 1935; Barlow, 1958, among others). This upper limit is decreased by increasing the area of the stimulus, and by increasing the background intensity. Barlow (1958) found that for a small stimulus on a low intensity background, temporal summation is complete up to 100 ms.

Simple integration may account for the time-luminance reciprocity described above, but there are other phenomena where it fails. For example, Ikeda (1965) found that when two flashes are presented in succession with a time separation of about 50 ms, the flashes do not summate positively to lower the visibility threshold. On the contrary, the combined threshold is higher than either separate threshold. This indicates that the temporal weighting function in summation is not constant, but varies with time having also negative values. A biphasic temporal impulse response function of the visual system can account for the results of double-flash experiments. This function has been discussed in greater detail in Section 2.3.2. It should be noted that temporal summation cannot be isolated from temporal impulse response - the result of temporal summation is always a weighted average, and the temporal impulse response gives the weights for this operation.

There is only one study in the literature that has concentrated solely on temporal summation of moving objects. Burr (1981) measured visibility thresholds for moving dots as a function of exposure duration. For speeds up to 16 deg/s, the thresholds were almost identical to those for stationary dots. In both cases, temporal summation was complete up to about 100 ms. For a dot moving at 16 deg/s, the spatial extent of summation is thus 1.6 degrees. This is more than twice the value Burr measured for a comparable stationary line, i.e. a line whose length was equal to the product of the dot speed and exposure duration, and whose luminance integral over time was the same as that for the moving dot. This cannot be explained by temporal and spatial integration, not even with biphasic temporal impulse responses.

In a related experiment, Burr (1980) measured the perceived amount of motion smear as a function of exposure duration. The measurements were made by matching the length of a short stationary line to the apparent length of dots in motion. The results are shown in Fig. 2.14. When the exposure duration was shorter than about 30 ms, the dots were seen to be smeared, the amount of smear approximating the distance the dot travelled during the exposure.

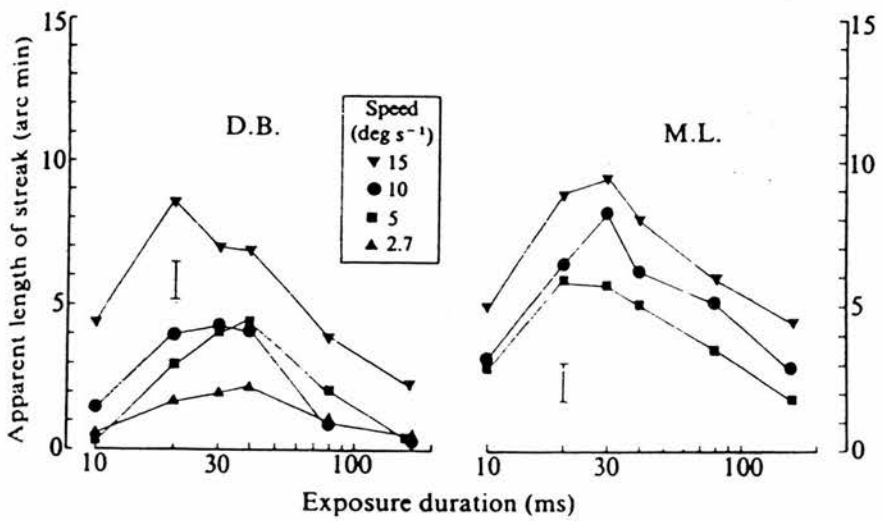


Fig. 2.14. The perceived smear for dots in a constant, rectilinear motion as a function of exposure duration. The amount of smear increases for durations up to 30 ms, and decreases after that. This is surprising, given that the visual system summates moving targets for over 100 ms. From Burr (1980).

However, for durations of longer than 30 ms, the perceived smear decreased with increasing durations. A simple temporal summation would have predicted a steady increase in smear up to the limit of complete temporal summation (about 100 ms), and a slight increase or an even level above the limit.

Burr noticed that at longer exposures, the impression of motion became much stronger. The amount of smear seemed to be connected to the strength of motion sensation - the smoother the sensation of motion, the sharper the dot was seen. Burr regarded this as an indication of that the motion detection mechanisms of visual system are also responsible for the analysis of form of moving objects, and that they remove smear if there is enough time for them to operate.

2.6. Models of motion deblurring

The temporal integration period of the visual system has been estimated to be over 100 ms for both stationary (Barlow, 1958) and moving objects (Burr, 1981; Burr, Ross & Morrone, 1986). One would expect from such a poor temporal resolution that motion blur would degrade our percepts of moving targets, or at least our thresholds for fine pattern discriminations involving them. However, we do not normally have any subjective impression of this blur (Burr, 1980). In addition, many hyperacuities are insensitive to motion up to velocities of several degrees per second (Westheimer & McKee, 1975; Westheimer & McKee, 1978; Morgan, Watt & McKee, 1983; Morgan & Benton, 1989).

Explanations for the absence of motion degradation in pattern discrimination tasks have consisted of two main approaches. First, it has been suggested (Burr, Ross & Morrone, 1986; Anderson & van Essen, 1987) that our visual system has special motion-deblurring or blur-prevention mechanisms to aid in the analysis of the shape of moving objects. In principle, these mechanisms restore positional acuity by taking into account the temporal delay at which different photoreceptors have been stimulated. The second approach (Morgan & Benton, 1989; Morgan, 1991) is against any general motion-deblurring mechanism. It assumes that image motion does introduce spatial blur. According to this approach, motion blur remains undetected because the internal representation of the object is already degraded by spatial blur to such a degree that motion blur does not noticeably further degrade the information.

2.6.1. Spatiotemporal receptive fields

Burr, Ross and Morrone (1986) suggested that the visual system contains motion detectors with receptive fields extended over space and time, and that these detectors analyse both the form and the motion of moving objects. They estimated the receptive fields in the following manner. First, they measured the minimum contrast required to detect the drift of a sinusoidal grating displayed together with phase-reversed masking gratings that varied over a wide range in both spatial and temporal frequency. From the masking results, contour plots of the spatiotemporal tuning functions were constructed. Fig. 2.15 shows these plots for one observer for four

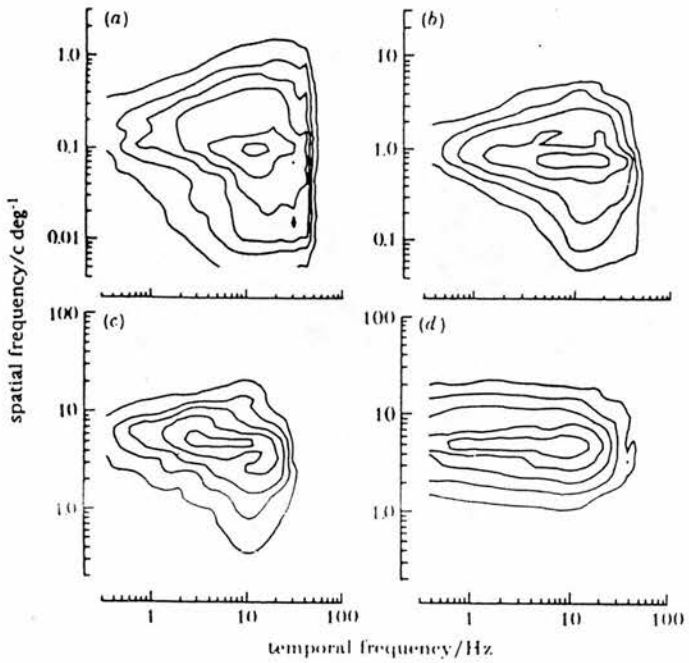


Fig. 2.15. Contour plots of the spatiotemporal tuning functions for four test gratings. The plots were constructed from the masking results using smoothing and interpolation with a two-dimensional cubic polynomial. Each contour line is separated by 0.5 log units. The test spatial and temporal frequencies (and drift velocities) were: (a) 0.1 c/deg, 8 Hz (80 deg/s), (b) 1 c/deg, 8 Hz (8 deg/s), (c) 5 c/deg, 8 Hz (1.6 deg/s) and (d) 5 c/deg, 0.3 Hz (0.06 deg/s). From Burr, Ross and Morrone (1986).

different velocities of the test grating. For test gratings drifting at 8 Hz, the tuning functions were all bandpass in both space and time, peaked at the spatial and temporal frequency of the test. For the practically stationary grating drifting at 0.3 Hz, the function was bandpass in space but lowpass in time.

To gain insight into the filter characteristics of motion detectors in the spatial and temporal domains, Burr et al. (1986) calculated the spatiotemporal weighting functions, or receptive fields, of the detectors with the inverse Fourier transform of the tuning functions. To simplify the demonstration, only the spatial dimension parallel to the direction of motion was considered. Since linearity is the basic requirement for the applicability of Fourier transform, Burr et al. assumed that the detectors behave linearly at threshold. Phase information is not given by the tuning functions,

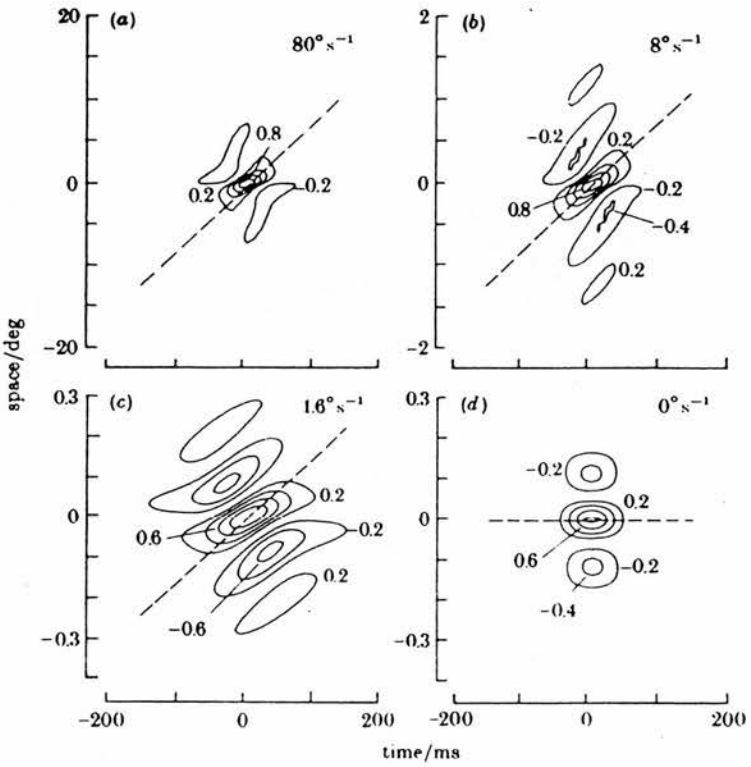


Fig. 2.16. Spatiotemporal receptive fields calculated by Fourier transform of the tuning functions of Fig. 2.15, assuming zero phase spectra. The dashed line in each field corresponds to the velocity of the test grating. From Burr et al. (1986).

and thus the phase spectrum had to be assumed. Based on some evidence from neurophysiological studies (Kulikowski & Bishop, 1981; Lee, Elephandt & Virsu, 1981), Burr et al. selected a zero phase spectrum, or linear phase, for their calculation. Fig. 2.16. shows the spatiotemporal receptive fields resulting from the transform of the tuning functions of Fig. 2.15.

All the receptive fields for motion-tuned detectors (Fig. 2.16 (a)-(c)) comprise parallel excitatory and inhibitory bands, elongated obliquely in space-time. Differently oriented fields correspond to different velocities of the test grating, and to different preferred velocities of the motion detectors. The field for the near-stationary case (Fig. 2.16. (d)) is not oriented in space-time, having a preferred velocity of zero deg/s. It has a

excitatory-inhibitory organization along the spatial axis, but the temporal profile is monophasic for any spatial location.

Burr et al. used the concept of spatiotemporal receptive fields to explain many phenomena of motion perception, one of which is the blur-free perception of moving objects. The essence of the explanation is simple: a spatiotemporally oriented receptive field integrates a moving object not statically, but along its path of motion, providing the extended temporal and spatial summation observed with moving objects (Burr, 1981). A range of detectors tuned to the velocity of an object but each with a different spatial frequency preference could cooperate to cancel the effect of motion and to allow the spatial analysis of the object as if it were stationary. Burr et al. (1986) proposed that the amount of motion smear will be determined by the width of the central region of the field, and by cooperation of many detectors. They claimed that "in principle definition as precise as desired may be obtained by the cooperative action of many fields of different cross-sectional profile." The cooperation would also play a major role in velocity coding.

The space-time model of Burr et al. (1986) is not the only model utilizing the principle of linear summation to produce motion tuning. Others include for instance the motion detection models of Adelson and Bergen (1985) and Watson and Ahumada (1985). While these models concentrate on the detection of different directions and speeds of motion, Burr et al. put more emphasis on the cooperative action of elementary motion detectors that may simultaneously detect both the form and the motion of an object.

2.6.2 Linear shifter circuits

Anderson and van Essen (1987) argued that in many computational visual problems it is desirable to control dynamically how the retinal output maps onto higher levels. One of their examples was the prevention of blurring of moving images. They proposed a general computational strategy for resolving these problems. Their solution, which they termed "shifter circuits", involves dynamic shifts in the relative alignment of input and output maps without loss of local spatial relationships. A schematic diagram of a simple shifter circuit is shown in Fig. 2.17. The shifts are produced in increments along a succession of relay stages connected by ascending, symmetrically diverging excitatory connections. At

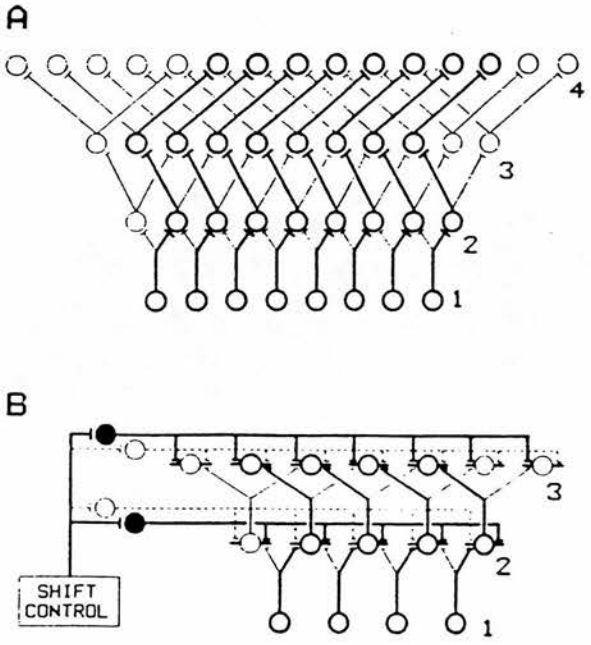


Fig. 2.17. A shifter circuit. (A) The ascending component of a four-layer circuit with eight cells at the bottom. Each cell connects to two target cells in the next level. (B) A complete three-layer circuit with four cells at the bottom. Connections for both ascending inputs and inhibitory neurones of shift control are shown. Active inhibitory neurones are represented by thick lines and filled cell bodies. Heavy ascending lines in A and B represent an activity pattern involving successive shifts to the right, left and in A, again to the right. From Anderson and van Essen (1987).

each stage the shifts are controlled by lateral inhibitory connections that selectively suppress appropriate sets of ascending inputs.

A descending visually driven feedback loop is needed to give a measure of the alignment to the shift control circuitry. In the case of motion blur prevention the feedback signals the locally measured retinal velocity. The aim is to transform a moving retinal image into a stationary neural image at the "stabilization stage". Knowledge of the retinal velocity is required to produce an appropriate cortical shift that compensates the effect of retinal motion. The strategy is illustrated in Fig. 2.18. The shifter circuit hypothesis has its strongest support in neurobiology:

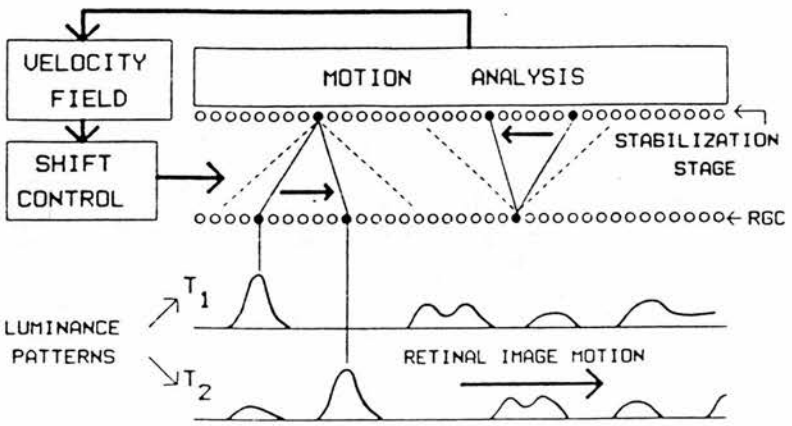


Fig. 2.18. A shifter circuit for motion compensation. A feedback from the retinal velocity measurement stage controls the shift mechanism to compensate for retinal image motion, and to form a stationary image at the cortical stabilization stage. From Anderson and van Essen (1987).

According to Anderson and van Essen (1987), the essential elements needed for assembling a shifter circuit, i.e. inhibitory interneurones and feedback, exist in the primate visual pathway.

The basic principle of both the shifter hypothesis and the spatiotemporal receptive field hypothesis is the same: integration follows the path of motion of the object. The implementation, however, may be different: if one assumes that spatiotemporal receptive fields are real physiological receptive fields, then they are stationary and made up of "hard-wired" connections. Shifter circuits, on the other hand, use dynamic switches and in fact shift whole receptive fields. However, if one regards spatiotemporal receptive fields as descriptive psychophysical receptive fields - like recently Anderson and Burr (1991) did, then the difference between the two mechanisms may not be significant.

2.6.3. A theory of no special mechanisms

Morgan and Benton (1989) showed that spatial interval discrimination thresholds rose about threefold over the velocity

range from zero to 6 deg/s when the gap separating the lines of the target was 4.5 min. This velocity dependence differs considerable from that found for vernier acuity (Westheimer & McKee, 1975; Morgan, Watt & McKee, 1983). Morgan and Benton took these different motion sensitivities of different hyperacuity tasks as evidence against general motion-deblurring mechanisms. According to their explanation, spatial interval acuity is more sensitive to motion blur than vernier acuity for the following reason. In the two-bar interval case, the bars follow one another rapidly at the same retinal location. Motion blur - or in other words the sluggish temporal resolution of the visual system - causes their responses to overlap, and their peaks become difficult to discriminate. On the other hand, however blurred the vertical vernier lines are by horizontal motion, they do not overlap on the retina. Morgan and Benton concluded that the insensitivity of vernier acuity to motion does not depend on the removal of motion blur, but on the fact that motion blur does not degrade the critical information in the stimulus.

3. Aims of the study

Humans perceive less motion blur than would be expected from the temporal summation of the visual system. The purpose of this work was to study why this happens, in particular whether there exist special mechanisms that prevent or remove motion blur, and what might be their nature. Both of the specific proposals presented in the literature (Burr & Ross, 1986; Anderson & van Essen, 1987) are based on motion-tuned mechanisms, but there are no experimental findings that would rule out other kinds of solution.

The main specific aims of the study were:

(a) to estimate the amount of equivalent intrinsic blur produced by motion as a function of velocity.

(b) to study the temporal characteristics of summation of objects in uniform rectilinear motion and of objects experiencing random spatial jitter.

(c) to evaluate the proposed motion-deblurring mechanisms on the basis of both theoretical considerations and the experimental data obtained.

(d) to search for plausible alternatives to motion-tuned models of motion perception and motion-deblurring.

4. Experiments on Gaussian blur discrimination as a function of blur and velocity

4.1. Introduction

The human visual system shows temporal integration over periods of the order of 100 ms (Barlow, 1958; Burr, 1981). This implies that all moving objects should appear blurred to a considerable degree. However, as we all know from our personal experience, this is not what we see. We are able to see moving objects with clarity or at least we are not conscious of motion blur. Some authors (Burr, Ross & Morrone, 1986; Anderson & van Essen, 1987) have suggested that there are general motion-deblurring mechanisms to produce the blur-free integration. Many very precise pattern discriminations are indeed unaffected by motion (Westheimer & McKee, 1975; Westheimer & McKee, 1978; Morgan, Watt & McKee, 1983; Welch & McKee, 1985; Morgan & Benton, 1989). On the other hand, Morgan and Benton (1989) found that spatial interval acuity is much more sensitive to motion than vernier acuity, and interpreted this as evidence against general motion-deblurring mechanisms. To put it simply, their explanation of the insensitivity of vernier acuity to motion is that precise vernier judgements can be made even if the target is spatially smeared - and it is the same whether this smearing is static or motion-induced.

However, having different motion sensitivities for different hyperacuity tasks is not necessarily a sufficient argument against a general motion-deblurring mechanism. Let us for instance assume that there is a motion-deblurring mechanism that removes most but not all of the motion blur predicted from the temporal summation period of the visual system. A mechanism removing all motion blur would not be biologically plausible - no neural mechanism can have exact temporal accuracy. Let us further assume that the residual motion blur is treated as equivalent to a small amount of spatial blur in the later stages of analysis. If we then have two hyperacuity tasks that have different sensitivities to spatial blur, it is easy to see that the tasks will also have different sensitivities to motion. Thus, having different motion sensitivities for different hyperacuity tasks is evidence against a

general motion-deblurring mechanism only if the sensitivities of these tasks to spatial blur can be shown to be equal. To study the issue properly one needs psychophysical tasks where the effects of static spatial blur and motion blur can be separated on a quantitative basis.

To decompose the effects of motion blur and static spatial blur one needs to investigate them using identical stimuli and psychophysical tasks. In the following experiments, blur discrimination for moving, Gaussian-blurred edges was examined. The basic experiment was a motion-domain extension of work by Watt and Morgan (1983b) who measured blur discrimination thresholds for stationary edges. Their results has been reviewed in Section 2.4.2 and in Fig. 2.8. In the present study, blur discrimination thresholds were measured as a function of both reference blur and velocity.

4.2. General methods

This section describes those experimental details that were kept unchanged through the whole series of blur discrimination experiments. Some setup parameters and stimulus conditions were chosen only after the preliminary experiments. These parameters and conditions will be described in connection with the main blur experiment. Any setup specific to a given experiment is described in connection with that experiment.

4.2.1. Apparatus

The stimuli were generated by a program running in a Macintosh IIcx computer. This program (excluding the adaptive probit estimation (Watt & Andrews, 1981) subroutines by Roger Watt) was written by the author in C language. In addition to the stimulus generation, the program also carried out most of the other tasks necessary for the execution of the experiment. To aid the validation of the results, it displayed the original data and the results of its analysis on the computer screen after each experimental run, and stored the data and the results together with the setup parameter values on a hard disk.

Stimuli were displayed on a Hewlett Packard model 1333a oscilloscope with a brief persistence P15 phosphor. The spot size

of this oscilloscope is less than 0.3 mm and the brightness of an intensified spot diminishes to 0.1% of its original value in less than 50 μs. Digital x and y position information for each dot were converted to analogue voltage levels by a DAC (digital-to-analogue converter) card. A long horizontal line was produced by generating a row of unresolved dots placed 0.3 min of arc apart. A 400 kHz triangle wave signal generated by a function generator was then added to the y signal to transform the horizontal line to a horizontal band. The band was moved by shifting its position less than 2.4 min of arc at least in every 5 ms so that the motion appeared to be continuous. The timing of movement and stimulus presentation was controlled by the VIA timers of the computer with an accuracy of about 5 μs. The minimum refresh rate was 400 Hz. The luminance profile of each band was controlled by a z modulation voltage provided by a third DAC channel. Luminance values of the display were calibrated against the input z voltages by a microphotometer.

4.2.2. Stimuli

Every band presented comprised a vertical edge of specific blur. A luminous vertical edge is a step change in luminance at a fixed horizontal coordinate. In mathematical terms an edge is a step function, and blurring an edge is defined by the convolution between the step function and the blurring function. The convolution of any given function with a step function is a special case of convolution: the result is the integral of that given function. In these experiments the blurring function was Gaussian, and thus the profile of the blurred edge had the form of an integrated Gaussian, i.e. the form of an error function. The blur width of an edge was specified by the standard deviation of the Gaussian. Fig. 4.1 illustrates the stimulus arrangement.

4.2.3. Procedure

Observers viewed the display from a distance of 100 cm. Viewing was binocular with natural pupils. The pupil diameter was not measured, but lighting conditions used in all final experiments of this thesis produce a diameter of about 4 mm (Watt and Morgan, 1983b). On each trial, the sequence of the patterns on the oscilloscope screen was the following: the fixation mark - a blank

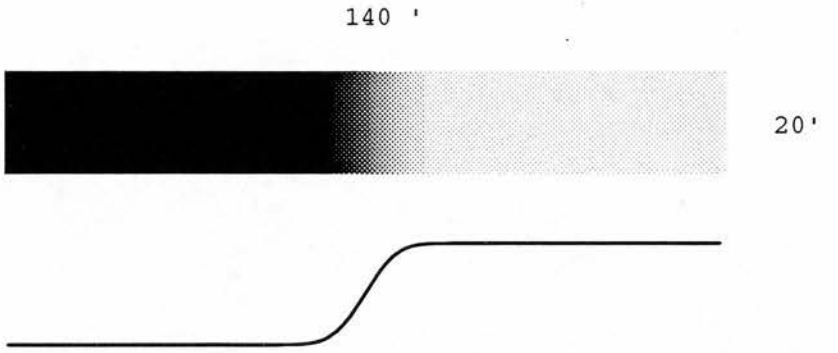


Fig. 4.1. The stimulus arrangement used in most of the blur discrimination experiments of this work. The luminous vertical edge has a profile of an integrated Gaussian, and this profile is shown below the band.

period - the first band - a blank period - the second band. The fixation mark was displayed for 500 ms. The bands were either stationary or moved horizontally on the screen randomly either leftward or rightward.

To minimize the effects of eye movements, the duration of the presentation of each band was 150 ms, known to be too short for the observer to track the object by smooth eye pursuit. One of the bands always had an edge with the reference blur width, and the other band had an edge with the test blur. The reference blur value was jittered from trial to trial in the range of the nominal reference blur ± 10%. Over a series of 64 trials an adaptive probit estimation (APE) algorithm (Watt & Andrews, 1981) was used to select the cue, i.e. the difference between the reference and the test blur, randomly from a number of preset magnitudes. The absolute value of this difference was always added to the reference blur value to get the test blur width. The sign of the difference was used to specify whether the band with the reference blur or with a test blur was presented first. The location of each edge within the band was varied randomly in a region of uncertainty of 2 arc min wide to make it impossible to the subject to use distance cues in the measurement of blur.

The observer's task was to decide whether the edge in the first or in the second band was more blurred. Threshold was defined as the standard deviation of the resultant psychometric function (83%

correct point) and it was estimated by fitting a cumulative normal curve to the psychometric function using probit analysis (Finney, 1971). Probit analysis also provides the standard error of the estimate for the standard deviation and a chi square value that can be used in assessing the goodness of fit. At least three independent thresholds were determined for each condition. Each final value represents the root mean square of these estimates. Each standard error presented is the standard error of mean of the estimates.

Three observers participated in one or more of the experiments. Observers RO (a graduate student) and MS (a research assistant) were naive as to the purpose of the study. Both had normal uncorrected vision. Observer AP (the author) participated in all experiments. He had uncorrected acuity of about a half of the normal mean value. Note that AP had no refractive errors - he is a slight binocular amblyope.

4.3. Pilot experiments

4.3.1. Experiment 1: the first pilot experiment

In this experiment thresholds for a difference in perceived extent of blur were measured as a function of blur extent and stimulus velocity. Five different reference blurs (0, 1, 2, 4, and 8 min of arc) and six different velocities (0, 1, 2, 4, 6, and 8 deg/s) were used. All the possible combinations of reference blur and velocity were measured in random order during one week.

Each band was 140 arc min in length and 20 arc min in height. The time interval between the presentation of the fixation mark and the first band was 500 ms. This was also the interval between the first and the second band. The experiment was conducted in dim room light. The minimum luminance in each band was always 2.9 cd/m^2 and the maximum was 23.0 cd/m^2, making a Michelson contrast of 77%. The background luminance on the screen was 0.8 cd/m^2. The bands in one trial had the same polarity of contrast, but two consecutive trials had different polarities, so that local retinal adaptation could be avoided.

The observer in the pilot experiments was the author (AP) who practised the task for several hours before starting to collect the data.

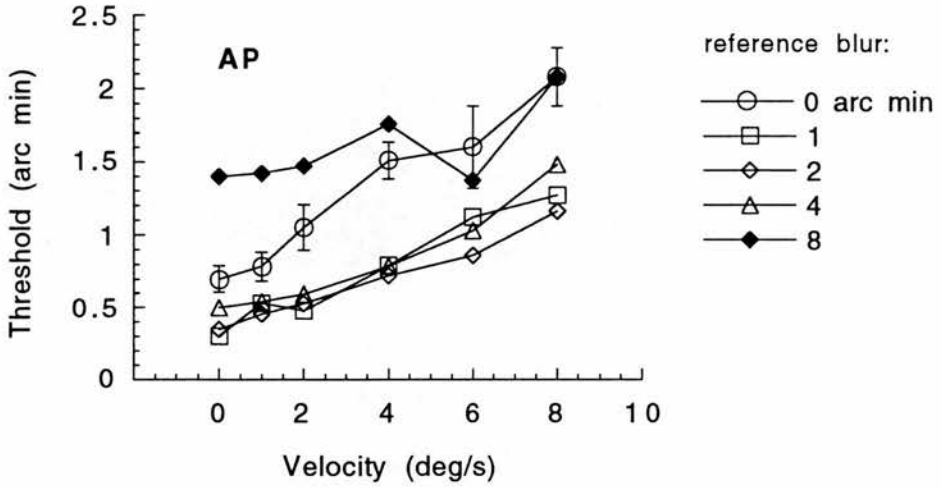


Fig. 4.2. Blur discrimination thresholds as a function of velocity at five different reference blurs (space constants 0, 1, 2, 4 and 8 arc min) for observer AP. The standard errors for zero arc min reference blur are presented as an example.

Results and discussion

Thresholds for difference in blur extent as function of velocity for five different criterion blurs are shown in Fig. 4.2. The basic finding is that the thresholds increase with velocity, indicating that motion does produce spatial blur. This increase is linear at least for reference blurs of 1, 2 and 4 arc min, and the rise is nearly fourfold over the range tested. The thresholds for these criterion blurs are also very close to each other at all velocities. To keep the figure easy to read, only the standard errors for a reference blur of zero arc min are presented. Typically one standard error varied in the range of 10 - 15 % of the threshold.

Fig. 4.3. shows discrimination thresholds as a function of reference blur for six different velocities. The thresholds depend on reference blur in a similar way for all velocities. Blur comparison is not at its best at zero blur but at some value between about 1 and 3 min of arc for all velocities. This finding is consistent with the result of Watt and Morgan (1983b) for a stationary edge (see Fig. 2.8). The value of the optimum blur does not show any clear relation to velocity.

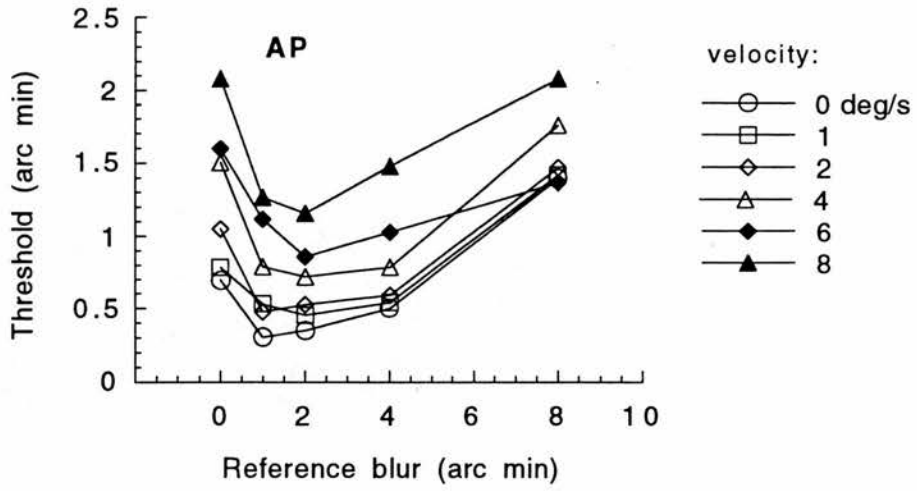


Fig. 4.3. Blur discrimination thresholds as a function of reference blur for six different velocities (0, 1, 2, 4, 6 and 8 deg/s) for observer AP.

The marked increase in thresholds with velocity argues against a perfect or a near-perfect motion-deblurring mechanism. On the other hand, the explanation for non-blurred percepts of moving objects based on a large degree of static internal blur predicts that the optimum blur should shift to higher blur values with velocity. This is because motion should just add a small amount of equivalent blur to the internal blur. No such shift is evident in Fig. 4.3. However, some observations on the original psychometric functions in this experiment suggested that the results might be erroneous, at least in some respects.

In many runs the measured data contained outliers, i.e. points that did not fit well to the psychometric function. An example of this is shown in Fig 4.4 (the figure also serves as an example of the display of results presented on the computer screen after each run). A true outlier does not belong to the probability distribution that defines the psychometric function, and that is why it should be removed before estimating the threshold. A true outlier may be due to a press on a wrong button after a correct decision in the comparison task, or to a wrong decision caused by a momentary drop in observer's concentration. The chi square value of the fit can be used to estimate whether a point is an outlier or not. If the removal of one point will decrease the chi square value markedly, say about 50% or more, that point is very likely to be

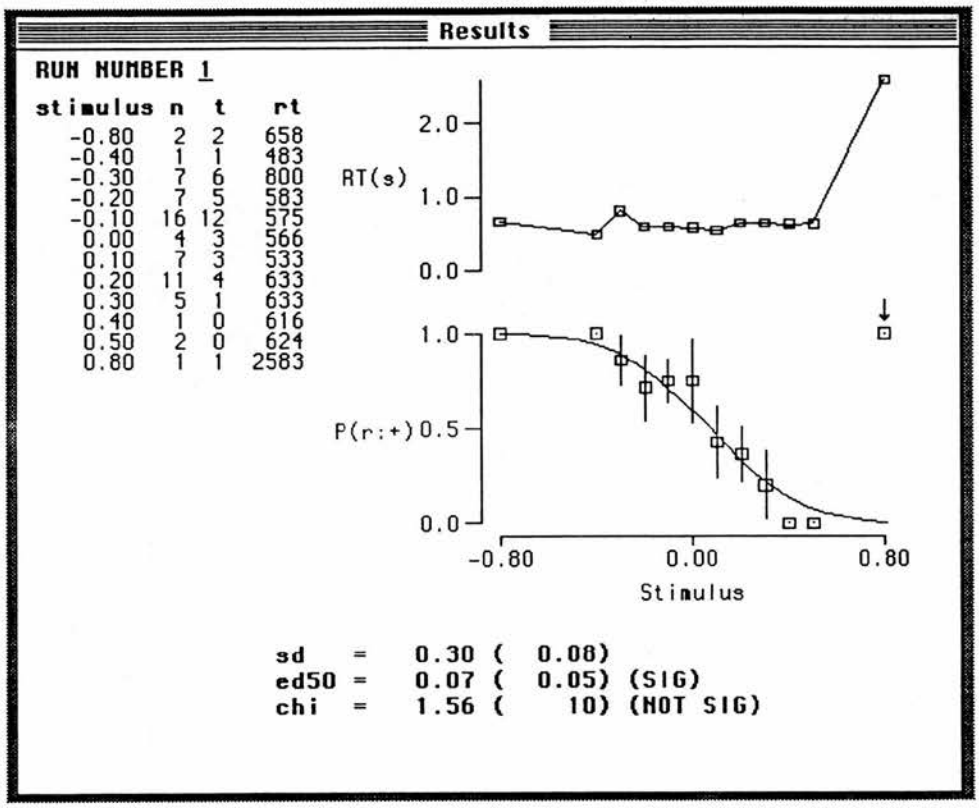


Fig. 4.4. An example of an outlier (pointed with an arrow) in the psychometric function. This figure is also an example of the results output presented on the computer screen after each run. In the figure, **stimulus** denotes the difference in the blur extent between the first and the second band in arc minutes, **n** denotes the number of trials at that difference, **t** denotes the number of trials where the observer reported that the second band was more blurred than the first band, and **rt** denotes the average reaction time in milliseconds. **sd** is the estimated threshold calculated as the standard deviation of the underlying Gaussian in the psychometric function, **ed50** is the bias, i.e. the deviation of the point of subjective equality from the point of physical equality (the standard errors of the threshold and the bias are presented in brackets), **chi** is a statistical measure of goodness-of-fit of the psychometric function, with the degrees of freedom of the fit in brackets. The upper graph in the figure plots the average reaction time, the lower graph plots the observed data and the fit to the psychometric function. In this figure, the fit is to the data from which the outlier has been removed. The ordinate of the lower graph denotes the probability of seeing the second band more blurred than the first band, with the data values obtained by dividing **t** by **n**.

an outlier, and it ought to be removed to keep the data valid. The same applies to the case of two suspected points but the removal of the points should be done under much greater consideration. If the data contain more than two points suspected to be outliers, the validity of the whole data is in question, and it is best if it is discarded. In this experiment, the psychometric data was edited according to these principles.

The chi square of a fit can be high even if there are no obvious outliers in the data. In this experiment the average ratio between chi square and degrees of freedom was 0.84. The value as high as this indicated that the psychometric functions were not as good as one would expect. One factor that could have degraded the psychometric functions was the existence of Mach-band-like illusory luminance bands at either sides of the edge. With some velocities and reference blurs, the observer could see an apparent low-luminance band at the foot of the edge, or an apparent high-luminance band at the upper end of the edge where the luminance gradient reached plateaux. Though resembling Mach bands, these bands were probably caused by the motion or the brief presentation of the edge together with the biphasic temporal impulse response of the visual system. To study how learning would affect the psychometric functions, the experiment was repeated with only slight modifications in the setup.

4.3.2. Experiment 2: the second pilot experiment

The main difference between this and the first pilot experiment was that in this experiment, the different combinations of reference blur and velocity were not measured in random order but all measurements for each reference blur were carried out before starting another. In this way the total variability of the blur extent in a session was reduced. It was expected that this might make it easier for the observer to learn to take the effects of Mach-band-like distortions into account. Another difference of course was that observer AP was now more experienced and better trained for the task than during the first experiment.

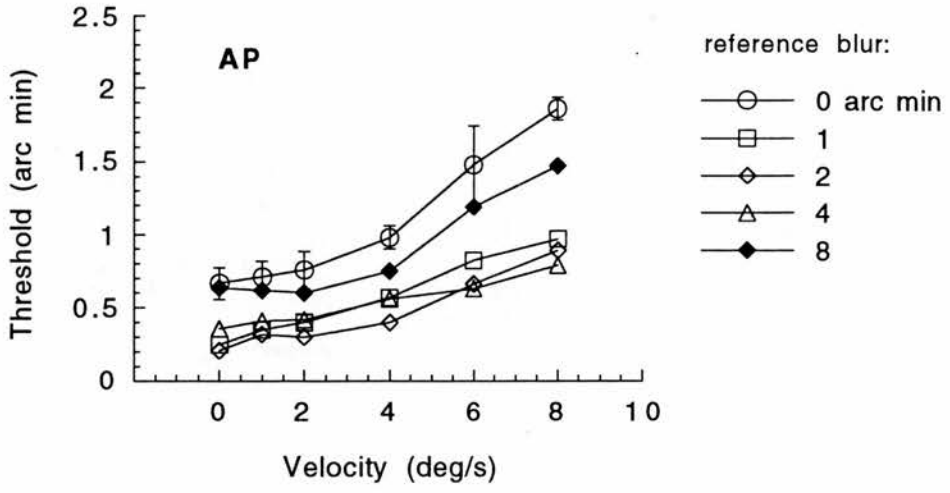


Fig. 4.5. Results of the second pilot experiment: Blur discrimination thresholds as a function of velocity at five different reference blurs for observer AP. The standard errors for zero arc min reference blur are presented as an example.

Results and discussion

Blur discrimination thresholds for five different criterion blurs are shown as function of velocity in Fig. 4.5. The same data are presented as a function of reference blur at six different velocities in Fig. 4.6. The basic findings are similar to those in the first pilot experiment. Thresholds increase with velocity, and the increase is fairly linear for all the reference blurs measured. The dependence on reference blur shows a U-shaped function for all velocities. This time the value of the optimum blur shifts towards higher blur values with velocity, consistent with the idea of equivalent intrinsic blur produced by motion.

On average, the standard errors were smaller in the second than in the first pilot experiment, being typically about 10% of the threshold. Also the chi square values were smaller, implying better fits to the probit function. However, there were still many outliers in the psychometric functions. To trace the origin of these outliers, some experimental runs were carried out with feedback, i.e. so that an incorrect response of the observer produced an auditory feedback tone. It was found that in some trials, the response was incorrect even if the observer was almost 100% certain

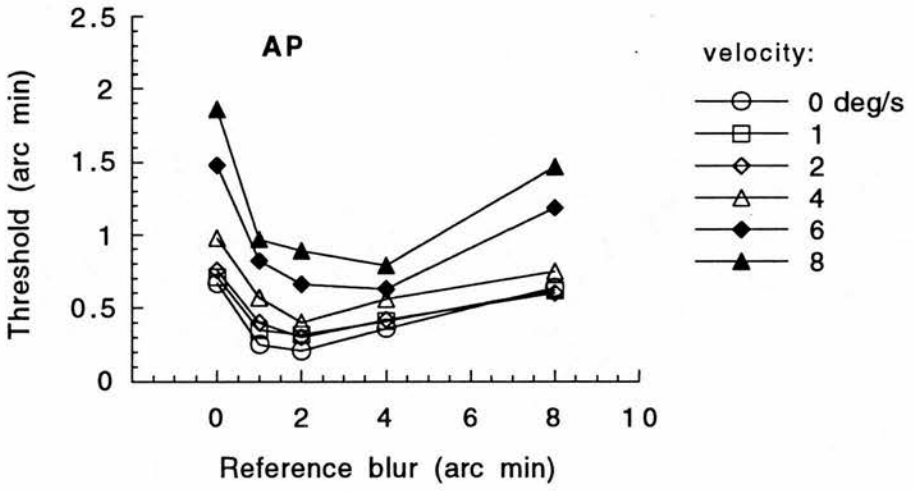


Fig. 4.6. Results of the second pilot experiment: Thresholds for blur discrimination as a function of reference blur for six different velocities (0, 1, 2, 4, 6 and 8 deg/s) for observer AP.

that he had given a correct response. A closer look at these trials revealed that outliers almost exclusively occurred when the first and the second band moved in opposite directions. Since the bands had the same polarity of contrast, this finding strongly suggested that the perceived blur extent of the edge moving with the low-luminance end in front was different from that of the edge having the high-luminance end leading.

4.4. Experiments on bias effects

The second pilot experiment showed that at least some of the poor fits of the psychometric functions could have been caused by a difference in the perceived blur extent between edges moving in opposite directions. In the pilot experiments, the contrast polarity was kept constant inside a trial in relation to the spatial coordinates. For instance, if the first band had the low-luminance end on the left, the second band also had the low-luminance band on the left. However, since the direction of motion of each band was random, the contrast polarity of the second band in relation to the direction of motion could be either the same or the opposite of that of the first band.

A difference in the perceived blur extent of bands having opposite leading edges can partly explain the large difference between the thresholds in the pilot experiment: In the first pilot experiment the series of measurements were carried out in random order, whereas in the second experiment all velocities for each reference blur were measured before starting another. If one assumes that the observer made his decision not just by comparing the first and the second edge but also by using information of his previous strategies in the same session, the less random presentation may have made it easier for the observer to learn to cancel out the effects of bias by mental arithmetic. Even without the change in the measurement order there might have been a large effect of learning - a task presumed to contain internal calculations as difficult as this is likely to show strong learning effects.

In a series of blur discrimination experiments, the effect of the direction of motion on the perceived blur was studied by measuring the bias, that is the deviation of the point of subjective equality from the point of physical equality, for different combinations of polarities between the first and the second band. The measurements were carried out by using a four-run version of an adaptive probit estimation procedure (Watt & Andrews, 1981). Every run consisted of 64 trials, making the total length of the session 256 trials. Each run corresponded to a particular combination of contrast polarities and directions of motion. For instance in run number three of the first bias experiment, the first band had the low-luminance end of the edge leading, whereas the second band with the same contrast

polarity had the high-luminance end leading after a change in the direction of motion. The runs were randomly interleaved to prevent the observer from anticipating the direction of motion of the second band on the basis of that of the first band. No feedback was given to the observer. This is a basic requirement of any bias measurement: the feedback would shift the point of subjective equality towards the point of physical equality, thus eliminating the bias.

4.4.1. Experiment 3: dim room light as adapting luminance

The first bias experiment was carried out in the same lighting conditions as the pilot experiments. The luminance in the dark end of each band was 2.9 cd/m^2 and in the bright end 22.9 cd/m^2, making a Michelson contrast of 77%. The background luminance on the screen was 0.8 cd/m^2. The standard deviation of the Gaussian for the reference blur used in this experiment was 4 arc min, and the velocity of the band was 4 deg/s. The four interleaved, different adaptive probit estimation runs consisted of those contrast-polarity and direction combinations that had been randomly mixed in every run of the pilot experiments. In the following text, the dark end refers to the low-luminance end of the edge and the bright end to the high-luminance end of the edge. The four interleaved runs were as follows: (1) in both the first and the second band, the dark end was leading, and the bands had the same direction of motion; (2) the bright end was leading in both bands, and the bands had the same direction of motion; (3) the dark end was leading in the first band, but the second band had an opposite direction of motion, and thus the bright end was leading; (4) identical to run number three but with dark and bright reversed. Observers AP and RO participated in this experiment.

Results and discussion

Biases for different conditions are presented in Fig. 4.7. Conditions A, B, C, and D correspond to runs (1), (2), (3) and (4) described above, respectively. When the first and the second band moved in the same direction and had the dark end leading (A), there was no bias in the perceived blur extent between the edges. When the bands moving in the same direction had the bright end leading (B), there was a small bias (about 0.2 arc min) in favour of seeing

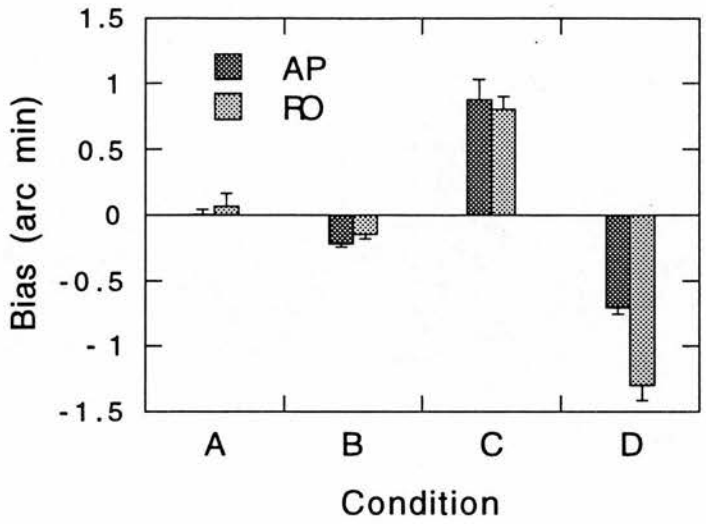


Fig. 4.7. Biases in the perceived blur extent of Gaussian blurred edges in dim room light for two observers (RO, AP) under four conditions: A. The dark end was leading both in the first and in the second band, both having the same direction of motion. B. Identical to A, except that the bright end was leading. C. The dark end was leading in the first band, but the second band had the opposite direction of motion and the bright end leading. D. Identical to C but with dark and bright reversed. The reference blur of the edge was 4 arc min and the velocity 4 deg/s. The bias is positive if the edge in the second band was seen more blurred. Error bars indicate one standard error. The width of the edge having the high-luminance end leading is perceived significantly larger than that of the edge with the low-luminance end in front.

the first band more blurred than the second. However, if the edges having the same contrast polarity moved in opposite directions (C,D), both observers perceived the edge having the high-luminance end leading about one arc min more blurred than the edge with the low luminance end in front, confirming the hypothesis that had inspired the experiment.

The large bias implies that the mechanisms for the perception of the form of moving objects are non-linear. It is easy to see that this nonlinearity must be temporal in nature. This is consistent with several findings from multiple-flash experiments (Kelly & Savoie, 1978; Watson, 1979; Bergen & Wilson, 1985) which have shown that the visual system is temporally nonlinear even at threshold. There are several possible explanations for this bias. The simplest one is based on the fact that a brief stimulus continues to be

visible for some time after its offset. This phenomenon - known as visible persistence - has been thoroughly reviewed by Coltheart (1980). However, this explanation is not as simple as it sounds. If visible persistence were to produce just a delay at the offset of the stimulus, it would not dilate the perceived blur extent. For this, a nonlinearity is needed. Such a non-linear property in visible persistence is the inverse intensity effect: the more intense the stimulus, the briefer its persistence. With a moving edge having the high-luminance end leading, the inverse intensity effect would increase the persistence of the low-luminance parts more than the high-luminance parts, making the perceived extent of the edge broader than its physical extent.

Another explanation for the bias is based on adaptation processes. These processes share at least part of their physiological background with visible persistence, but the terminology is different. For foveal cone vision, adaptation mechanisms can be divided into two classes: multiplicative mechanisms which reduce the gain and subtractive mechanisms which filter out the background signal (Hayhoe, Benimoff & Hood, 1987; Hayhoe, 1990). Hayhoe et al. (1987) showed that, at the onset of an adapting light, the multiplicative change is accomplished within about 50 ms. Also most of the subtractive change occurs rapidly. At the offset of an adapting field, the multiplicative process takes over 200 ms to recover. This difference in the time-course of the multiplicative process between onset and offset could explain why the edge having the high-luminance end leading (mimicking offset) is perceived significantly broader than the edge with the high-luminance end trailing (mimicking onset).

A third explanation is based on the possibility that the ON and OFF pathways of the visual system may produce asymmetric contributions to the perception of the form of moving edges. Some psychophysical studies have indirectly shown that these pathways may be separable in their response to adaptation. For instance, Moulden and Begg (1986) demonstrated the existence of adaptation which was specific not only to the direction of motion of an edge, but also to its contrast polarity. Recently, evoked potential studies have shown an asymmetric response to light increments and decrements (Zemon, Gordon & Welch, 1988). One single mechanism based on linear concepts cannot explain the bias found in this experiment. However, the results do not rule out the possibility that two or more separate mechanisms, such as for instance ON and OFF systems, could

explain the bias, for instance on the basis of their different temporal impulse responses.

4.4.2. Experiment 4: normal room light as adapting luminance

The bias experiment in dim room light showed a difference in the perceived blur extent between edges having opposite contrast polarities leading in motion. This experiment was conducted in normal room light in order to see if an increase in adapting luminance would reduce the bias effects. Adapting luminance is known to affect the temporal behaviour of the visual system in such a way that in general, visual processing happens faster in the light. Most of the evidence has come from temporal impulse response studies (Uetsuki & Ikeda, 1970; Kelly, 1971; Ueno, 1977, among others), but there is also evidence from neurophysiology; for example, Whitten and Brown (1973) showed that photoreceptor responses decay far more rapidly in the light than in the dark. Recently, Dawson and Di Lollo (1990) reported a strong effect of adapting luminance on the temporal limits of motion perception: decrements in luminance produced significant increments in the maximum interstimulus interval at which coherent motion was perceived in a two-frame random dot experiment.

The setup of this experiment was the same as in the previous one, except for the luminances of the band and the background. The luminance in the dark end of each band was now 28.5 cd/m^2 and in the bright end 58.9 cd/m^2. The background luminance on the screen was 25 cd/m^2. These luminance values were a result of using normal fluorescent roof lighting in the experiment room, and multiplying the absolute luminance values used in the previous experiments by a factor of 1.5. This multiplication was done to keep the visual appearance of the band similar to that in the previous experiments. However, contrast was still reduced from 77% to 35%. On the other hand, many studies have shown that the characteristics of the motion system are independent of contrast above a certain critical value, and in all these studies this value has been well below 35 % (Nakayama & Silverman, 1985; McKee, Silverman & Nakayama, 1986; van de Grind, Koenderink & van Doorn, 1987, among others). For instance, Nakayama and Silverman (1985) derived a nonlinear contrast response function for the motion system and found this to saturate in the neighbourhood of 2 to 3 %.

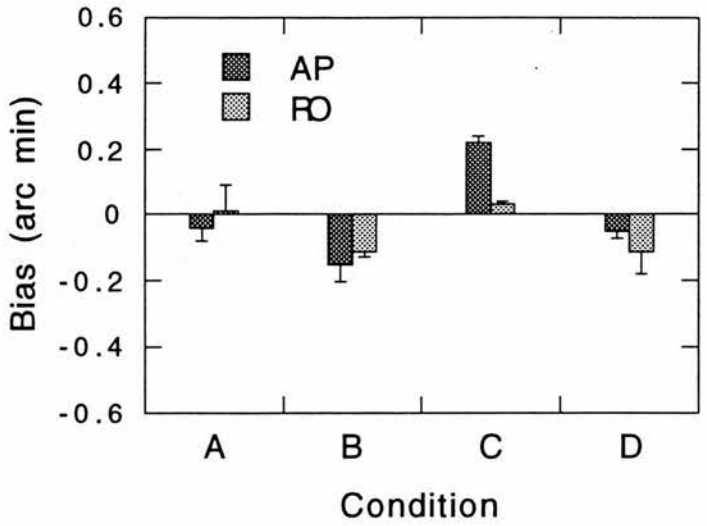


Fig. 4.8. Biases in the perceived blur extent of Gaussian blurred edges in normal room light for two observers (RO, AP) and four conditions, which are the same as in Fig. 4.7. The reference blur of the edge was 4 arc min and the velocity 4 deg/s. The bias is positive if the edge in the second band was seen more blurred. Error bars indicate one standard error. The edge having the high-luminance end leading is perceived broader than the edge with the low-luminance end in front (conditions C and D), but on the average the difference is only about 0.1 arc min.

Results and discussion

Biases for different conditions are presented in Fig. 4.8. The conditions are described in connection with the bias experiment in dim room light and in the text of Fig. 4.7. When the first and the second band moved in the same direction and had the dark end leading (condition A), there was no bias in the perceived blur extent between the edges. Condition B where the bands moved in the same direction with the bright end leading produced a small bias (the mean value of the two observers was 0.13 arc min) of seeing the first band more blurred than the second. The biases in conditions A and B were close to the corresponding values in dim room light. When the edges had the same contrast polarity but moved in opposite directions (C and D), there was a small but consistent bias (on the average about 0.1 arc min) of seeing the edge with the high-luminance end leading more blurred than the edge with the low luminance end in front. Thus, the direction of bias was the same but the amount of it was reduced to about one tenth of that found in

the experiment conducted in dim room light. The results confirmed that an increase in adapting luminance reduces the bias effects.

4.4.3. Experiment 5: randomized time intervals in trials

The previous experiment showed that by changing adapting luminance, the bias effects can be reduced. However, they cannot be totally removed without removing those trials where the first and the second band have opposite contrast polarities leading in motion. This cannot be done by simply removing conditions C and D used in the previous experiments. If there were only conditions A and B present, the observer could predict the direction of the second band after seeing the first and would be able to make anticipatory eye movements which might help him to track the band. In this experiment, conditions C and D were changed as follows: C. The first band had the dark end leading. In the second band both the contrast polarity and the direction of motion were changed, and thus it too had the dark end leading. D. Analogous to condition C, except that bright end was leading in both bands.

In the setup of all the previous experiments there has been a factor that could have aided the observer to produce subconscious anticipatory eye movements to track the target. This factor is the constant time interval of 500 ms that was used between the presentations of the fixation mark and the first band, and also between the first and the second band. This constancy was removed from this and all the rest experiments of this thesis by randomizing the temporal intervals. At each trial, the time interval between the presentations of the fixation mark and the first band was selected randomly from a uniform probability distribution extending from 300 ms to 500 ms. A similar randomization was applied to the interval between the presentations of the bands but in the range of 450 to 750 ms. All the other setup values were kept the same as in the bias experiment conducted in normal room light. The observer was the author.

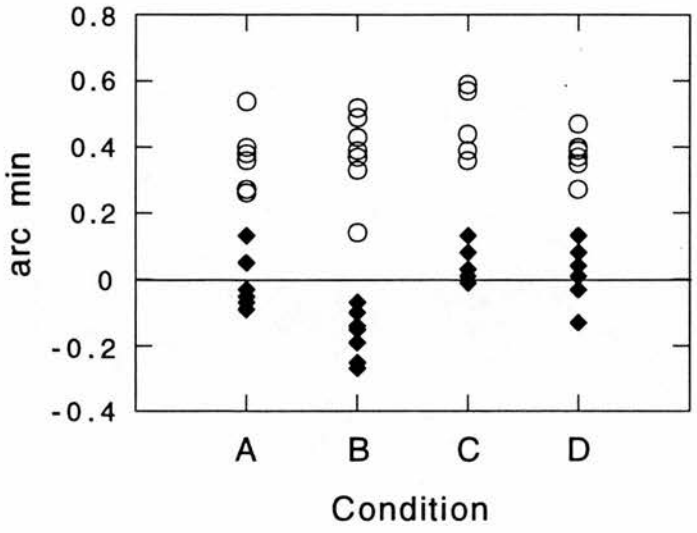


Fig. 4.9. Biases (filled diamonds) and thresholds (open circles) for the perceived blur extent of Gaussian blurred edges for observer AP. The experiments were carried out in normal room light with randomized time intervals between the presentations of the fixation mark and the bands. The reference blur of the edge was 4 arc min and the velocity 4 deg/s. Results from six independent measurements are shown to clarify the variability of the data. In each of the four conditions, the first and the second band had the same leading end. In conditions A (dark leading) and B (bright leading), the bands had the same direction of motion. In conditions C (dark leading) and D (bright leading), the bands had opposite directions of motion. The bias is positive if the edge in the second band was seen more blurred. The biases are very close to zero except for condition B which has a mean bias of 0.17 arc min.

Results and discussion

To clarify the variability of the data, all individual data points for biases and thresholds are presented in Fig. 4.9. The biases are practically zero for all conditions except for condition B where the bright end was leading in both bands and the bands moved in the same direction. This condition has a mean bias of 0.17 arc min. The mean values for the thresholds are in the range of 0.38 and 0.47 arc min.

The data shows that biases can be brought very close to zero by letting the bands move with the same end leading and by minimizing the possibility of anticipatory eye movements by randomized time intervals. The only exception is the condition where the bright end leads in both bands and the bands move in the same direction. This condition has through all the bias experiments shown a consistent

bias of about 0.15 arc min of seeing the first band more blurred than the second. A bias of this direction would result in if the observer were using a strategy where he would try to track the second band by assuming that it will move in the same direction as the first band. However, this explanation can be ruled out by the fact that this strategy would also predict an opposite bias in condition D, and D has no such bias. Thus, the explanation of this bias is most likely to be adaptation: the perceived width of the second band is affected by the adaptational changes the first band generates. Since the bias exists only in condition B and not in D, this adaptation is specific both to the direction of motion and to the contrast polarity of the edge. Since there is no discernible bias in condition A, this adaptation seems to have effect only in the case which mimics the offset of an adapting field.

Moulden and Begg (1986) demonstrated a simultaneous specificity of adaptation to the direction of motion and the contrast polarity of an edge. However in their experiment, the adaptation period was 60 s in duration. This experiment demonstrates adaptation effects of the similar specificity in a much shorter time scale. To the author's knowledge, no such adaptation effects have been reported in the literature.

The results of the bias experiments were used to help to choose the conditions of contrast polarities and directions of motion for the rest of the blur discrimination experiments. The basic requirement was that the observer must not be able to conclude the direction of motion of the second band on the basis of the contrast polarity and the direction of motion of the first band. Condition B was discarded because of its bias, and thus condition A had to be selected. This was not a bad selection - condition A had practically a zero bias in all the bias experiments. Condition C was selected to be the condition that contains opposite directions of motion. Condition D could not be selected because in the combination of A and D, the observer would have been able to predict accurately the direction of motion of the second band in each trial. Conditions A and C were not mixed in one adaptive probit estimation run, instead they were randomly interleaved into two runs that were analyzed separately.

4.5. Experiment 6: main experiment

To separate the effects of motion blur and static spatial blur, blur discrimination for moving, Gaussian-blurred edges was examined in this experiment. Blur discrimination thresholds were measured as a function of both reference blur and velocity. A more detailed introduction to the purpose of blur discrimination experiments of this thesis was given in Section 4.1.

4.5.1. Methods

Most of the stimulus arrangements and setup parameter values of this experiment were the same as those in the pilot experiments (see Section 4.2). The results of the bias experiments, however, led to some significant adjustments. The adjusted parameter values and conditions are described below.

In this experiment, normal room lighting provided a veiling luminance of 25 cd/m^2, ensuring that all observations were carried out at photopic levels. The minimum luminance of the band was always 28.5 cd/m^2 and the maximum 58.9 cd/m^2, making a Michelson contrast of 35%. On each trial, the sequence of visual patterns on the oscilloscope screen was the following: the fixation mark - a blank period - the first band - a blank period - the second band. To minimize the effects of anticipatory eye movements, the durations of the blank intervals were randomized, the first in the range of 300 - 500 ms and the second in the range of 450 - 750 ms. The bands were either stationary or moved horizontally with the same speed randomly either leftward or rightward. The polarity of the edge was selected so that the low luminance end of the band was always leading. Two series of 64 trials corresponding to situations where the bands moved either in the same or in the opposite directions were randomly interleaved. The analysis of the resultant two psychometric functions was done separately. At least four thresholds were determined under each condition of reference blur and velocity. Each final value reported represents the root mean square of these estimates. Three observers (RO, MS and AP) measured their blur discrimination thresholds for all the possible combinations of four different reference blurs (0, 1, 2, and 4 min of arc) and six different velocities (0, 1, 2, 4, 6, and 8 deg/s).

In addition, observer AP also measured thresholds at a reference blur of 8 arc min.

4.5.2. Results

Fig. 4.10 shows blur discrimination thresholds as a function of velocity with four different reference blurs for the three observers. Although the data of observer AP differ in some respects from those of the others, the main features are similar. For each reference blur the thresholds increase approximately linearly with velocity, and the slope of this increase is inversely related to the reference blur. The smaller the reference blur, the larger is the effect of velocity on the thresholds. Blur comparison is not at its best at zero blur but at some higher reference blur value for all velocities. This is consistent with the finding of Watt and Morgan (1983b) for stationary Gaussian blur. This optimum blur also seems to shift to higher blur values with velocity. Observer AP's performance is better than that of the other observers at a reference blur of 4 arc min, whereas the others perform better than him at smaller reference blurs. The thresholds for observer AP at a reference blur of 8 arc min are shown later in connection with the model fitting of the data.

4.5.3. Models of blur discrimination of moving targets

The results show that image motion shifts blur discrimination thresholds, indicating that motion produces equivalent spatial blur. To estimate the amount of this equivalent blur, we need a model to separate the effects of motion blur and static spatial blur. Using a linear mathematical model has a prerequisite: we must assume that the blur discrimination system is linear near threshold. Another fact of signal analysis helps us in building the model: in mathematical terms blurring is equivalent to filtering since both are produced by convolving the signal with a function that can be termed either a blurring function or a spatial filter. In filtering terminology, the results of this experiment indicate an increase in the effective spatial filter size with velocity.

The spatial filtering of moving stimuli must contain at least some of the same static spatial filtering operations that blur stationary images since there is no difference in the optical blurring of stationary and moving objects. Nor does the mosaic of

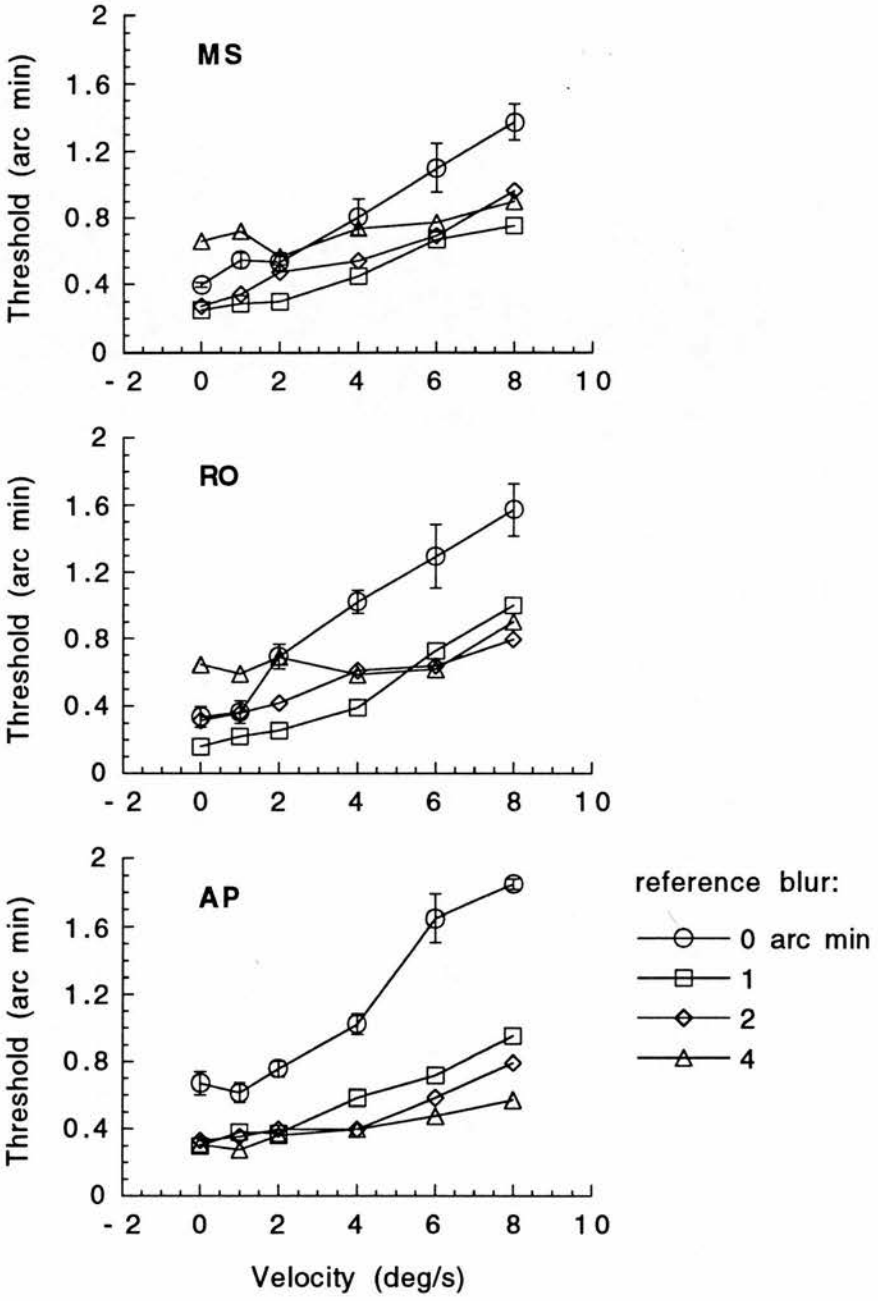


Fig. 4.10. Blur discrimination thresholds as a function of velocity at four different reference blurs (space constants 0, 1, 2 and 4 arc min) for three observers (MS, RO and AP). The reference blur space constant is specified by the standard deviation of the Gaussian. The error bars for zero arc min reference blur are presented as an example, they represent ± one standard error. Typically one standard error was about 10 % of the threshold.

photoreceptors change in response to moving images. It is assumed here that all the operations that are not velocity dependent are actually the same, and thus represent the compound action of these operations with a single static spatial filter of space constant s. In addition, the model has a blur component which is velocity dependent. In the simplest case the space constant of this component is the product of a parameter f and a velocity v. The way to combine the static and velocity dependent spatial filters, and the interpretation of the parameter f and the product fv depends on the assumed motion blur type. To clarify this, two simplified examples of different motion blur types are given.

If motion blur is purely camera-like, the filtering can be described in two phases. First, the motion of the object and the temporal response of the photoreceptors produce together the velocity-dependent spatial filtering, and the outputs of the receptors are then combined by the static spatial filter. Mathematically the operation contains two separate, consecutive convolutions, and the total output is separable in terms of static and velocity-dependent spatial filtering. Since the variances are additive in convolution, the space constant of the resultant spatial filter is the square root of the sum of the squared space constants of the components. In the case of camera-like blur, the parameter f is a measure of the exposure or summation time; in fact it is the standard deviation of the temporal weighting function, or the temporal impulse response of the system.

Another type of motion blurring results if we assume that the visual system uses different spatial filters at different velocities, and that the size of these spatial filters increases with velocity. If, for the simplicity of this example, we ignore the temporal response of the photoreceptors, then the filtering operation in this case actually contains only one convolution. The total output is inseparable in terms of static and velocity-dependent spatial filtering, and the space constant of the filter is simply the sum of the components. The velocity-dependent component fv is not a space constant of any realisable filter, it just indicates how the size of the spatial filter depends on the velocity of the object. The parameter s gives the space constant of the filter for stationary objects.

In reality, blurring of moving objects is likely to be a rather complex mixture of separable and inseparable blur components. There

is no way to avoid optical blurring that produces in principle a static filtering stage. Furthermore, it is hard to believe that the sluggish temporal response of photoreceptors would not produce any motion blur on moving objects. The static blur and the velocity dependent blur produced at these two stages are separable. On the other hand, at the bipolar or ganglion cell level, the temporal filtering in the centre of a receptive field is different from the temporal filtering of the surround (Barlow & Levick, 1969a; Barlow & Levick, 1969b; Derrington & Lennie, 1984; Fleet, Hallett & Jepson, 1985). Thus, the static and the velocity dependent blurs produced by the centre-surround organization are inseparable. There may also be some higher filtering stages producing blur of either or both types. In spite of all this complexity, the data can, however, be modelled by using only one static and one velocity-dependent spatial filtering component. The estimates obtained this way are the total effective space constants of the components. If we calculate the estimates for two extreme cases - fully separable and fully inseparable blur - we get the lower and upper limits for the effective values.

This model derivation follows some of the reasoning presented by Watt (1988) for the discrimination of stationary blur. The internal representation of an edge is a blurred image of the real physical appearance of that edge. In the case of fully separable blur, it is assumed to be blurred by the two spatial filters: a static one and a velocity-dependent one. In mathematical terms, the representation of the blurred edge is determined by the convolution of the real edge of blur B, a static spatial filter with space constant s and a velocity-dependent spatial filter with space constant fv. The standard deviation or the blur width of the internal representation of the edge B' is the square root of the sum of variances of its convolution components:

$$B' = \sqrt{B^2 + s^2 + (fv)^2} \tag{4.1}$$

Let us then suppose that there is only one dominant source of error in the judgement of blur difference. This source is the comparison of measurements of internal representations of blurred edges, which introduces a Weber's Law error:

$$\Delta B' = kB' \tag{4.2}$$

where k is the Weber fraction and $\Delta B'$ the threshold in the comparison of internal representations of edges. If we then assume that the internal difference $\Delta B'$ is caused by the increment ΔB to the reference blur, we have

$$B' + \Delta B' = \sqrt{(B + \Delta B)^2 + s^2 + (fv)^2} \tag{4.3}$$

Substituting Eq. (4.2) leads to

$$(1 + k)B' = \sqrt{(B + \Delta B)^2 + s^2 + (fv)^2} \tag{4.4}$$

Substituting Eq. (4.1), we have

$$(1 + k)\sqrt{B^2 + s^2 + (fv)^2} = \sqrt{(B + \Delta B)^2 + s^2 + (fv)^2} \tag{4.5}$$

Squaring both sides, cancelling equal terms and reorganizing terms, we obtain

$$\Delta B^2 + 2\Delta B - (k^2 + 2k)[B^2 + s^2 + (fv)^2] = 0 \tag{4.6}$$

Solving Eq. (4.6) for ΔB gives us

$$\Delta B = -B + \sqrt{B^2 + (k^2 + 2k)[B^2 + s^2 + (fv)^2]} \tag{4.7}$$

Equation (4.7) is the model for blur discrimination threshold as a function of reference blur and velocity for fully separable blur.

In the case of fully inseparable blurs, the sum of the static spatial and the velocity-dependent component is treated as one filter, resulting in a slightly different model:

$$\Delta B = -B + \sqrt{B^2 + (k^2 + 2k)[B^2 + (s + fv)^2]} \tag{4.8}$$

If our visual system really uses special spatial filters for moving objects, there are no particular reasons why the size of the spatial filter should be a linear function of velocity. The filter size could, for instance, increase faster than the linear relation

predicts. The simplest case of this kind of behaviour is a quadratic velocity dependence, which gives us the following model:

$$\Delta B = -B + \sqrt{B^2 + (k^2 + 2k)[B^2 + (s + fv + gv^2)^2]} \qquad (4.9)$$

Camera-like blur can produce only a linear velocity dependence, and so if this quadratic model fits to the data significantly better that the model with a linear velocity dependence, we can consider it as evidence for the existence of special spatial filters for moving objects.

4.5.4. Fitting the models

A standard nonlinear least-squares routine, the Levenberg-Marquart method (Press & William, 1988), was applied to fit the models described by Eqs. (4.7), (4.8) and (4.9) to the data. The maximum likelihood estimates of the model parameters were obtained by minimizing the χ^2-merit function, which was defined as follows:

$$\chi^2(s,f,k) = \sum_{i=1}^{N} \left[\frac{\Delta B_i - \Delta B(B,v;s,f,k)}{\sigma_i} \right]^2 \qquad (4.10)$$

where $\Delta B(B,v;s,f,k)$ is the model evaluation at reference blur B, velocity v and parameter set $\{s,f,k\}$, and σ_i is the standard error for the measured value ΔB_i. In addition to model parameters, the fitting procedure provided error estimates for the parameters and a statistical measure of goodness-of-fit. The probability that the true error of each parameter is less than the given estimate is 68%, assuming that the errors are normally distributed. The goodness-of-fit of a χ^2-fit is a function of the minimized χ^2-value and the degrees of freedom of the fit. The quantitative absolute value of the goodness-of-fit is the probability Q that the χ^2-value obtained should result from chance fluctuations of the data. Both the χ^2-value and the probability Q depend on the estimates of the measurement errors. If the probability Q is very small, then the discrepancies from the model are unlikely to be chance fluctuations: either the model is wrong, or the measurement errors are larger than estimated. However, due to the fairly common situation of a non-

normal distribution of measurement errors, models with Q-values as low as 0.001 are often deemed acceptable (Press & William, 1988).

4.5.5. Results from the fitting

With all the models and measurement data, the method was found to converge in a few iterations to the best fit parameter set. This happened even when the initial guess values for the parameters were far away from the best fit set, indicating that the minima found by the method are the true global ones. In all cases, this was confirmed by graphical inspection of the merit function, which also ruled out the existence of any other significant local minimum, i.e. the merit functions were found to be unimodal, with a single minimum.

Fig. 4.11 shows the fit of the fully inseparable blur model with a linear velocity dependence superimposed on the original data. The model fits to the data well for each observer, except for observer AP in the case of the sharp edge (a reference blur of 0 arc min). The quantitative values for goodness-of-fit, i.e. the Q-values are presented in Table 4.1. The number of independent data points is twenty-four for each observer, and thus the number of degrees of freedom is twenty-one for the models with three parameters (Eq. (4.7) and (4.8)) and twenty for the model with four parameters (Eq. (4.9)). A rule of thumb in chi-square fitting states that a typical value of χ^2 for a moderately good fit is approximately equal to the degrees of freedom of the fit, and for observer RO, the χ^2-values are close to that. The fits for observer MS are not as good, but the Q-values are still several orders of magnitude higher than 0.001. For observer AP, the overall fits are much poorer than for the other observers, but the fits are not uniformly poor across the reference blurs. Approximately two thirds of the total χ^2-value originates from the reference blur of zero arc min. In fact, when the data for zero arc min were excluded from the fits (AP-0' in Table 4.1), the Q-values rose to several percents, i.e. to values not far from those for observer RO. It is unclear why the fit to the sharp edge data of observer AP is so poor. One reason might be that observer AP practised the task much more than the other two observers, and most of this practising was done at small reference blurs different from zero and with low velocities. Learning specific to a certain velocity and/or reference blur region could have changed the performance in a nonuniform way.

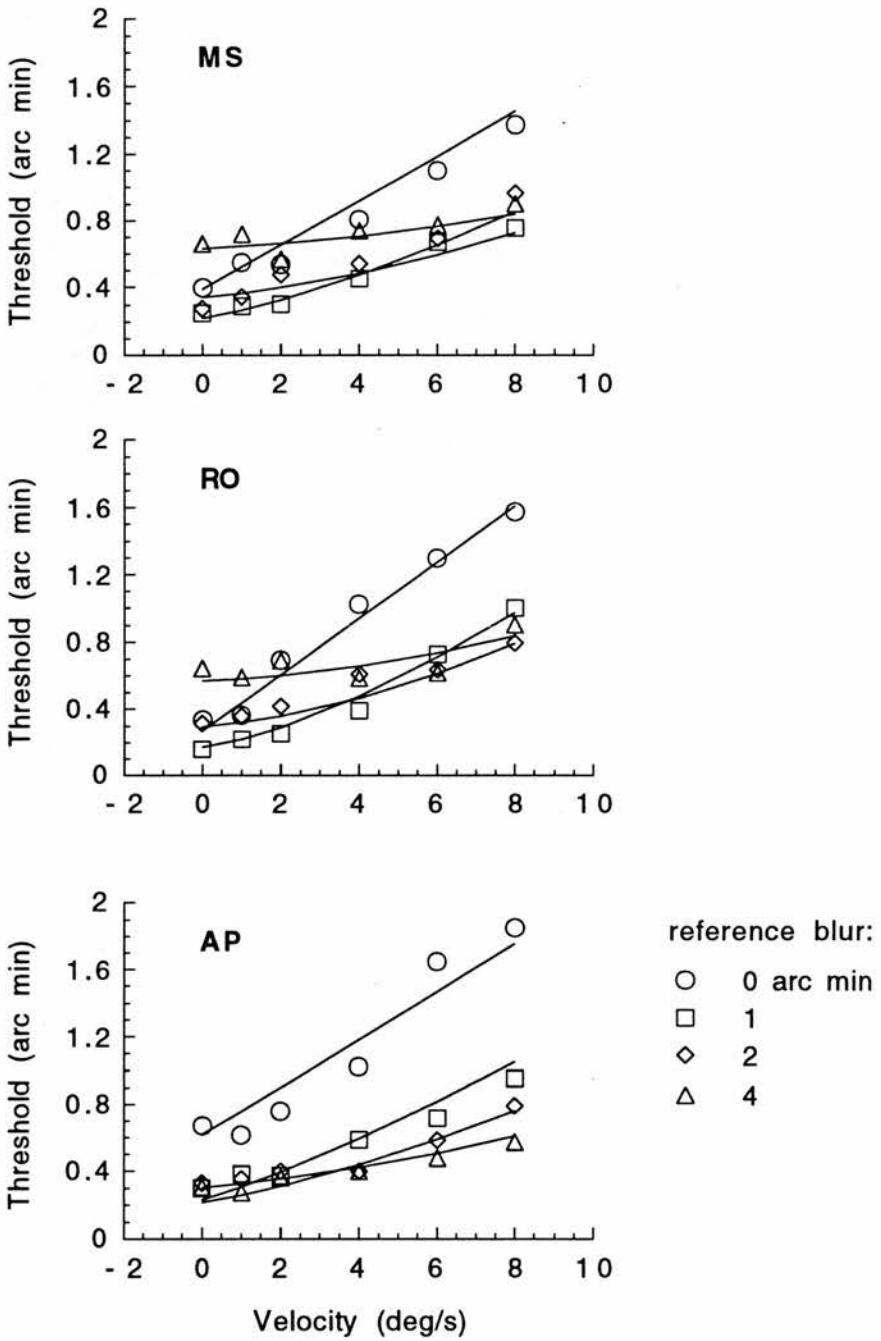


Fig. 4.11. A blur model fit superimposed on the original blur discrimination threshold data from Fig. 4.10. The model employed here assumes that the static and the velocity-dependent component of the effective spatial filter are fully inseparable, and that the velocity dependence is linear.

Table 4.1. Modelling of blur discrimination thresholds: best fit parameters, their standard errors, χ^2- and Q-values for three different models (labelled according to the equation numbers in text: (4.7) fully separable blur, (4.8) fully inseparable blur with a linear velocity dependence, and (4.9) fully inseparable blur with a quadratic velocity dependence) and three different observers (MS, RO, AP). For observer AP, the data were modelled also after excluding the thresholds for a reference blur of zero arc min (labelled AP-0').

	s	σs	f	σf	g	σg	k	σk	χ^2	Q
MS										
(4.7)	0.71	0.03	5.41	0.21			0.155	0.004	44.03	0.23
(4.8)	0.68	0.03	3.84	0.19			0.154	0.004	43.76	0.25
(4.9)	0.69	0.03	2.90	0.54	0.14	0.08	0.155	0.004	40.28	0.46
RO										
(4.7)	0.63	0.07	6.40	0.33			0.140	0.006	26.11	20.21
(4.8)	0.49	0.06	5.10	0.35			0.140	0.006	22.36	37.92
(4.9)	0.51	0.08	4.76	0.95	0.05	0.13	0.140	0.006	22.19	33.01
AP										
(4.7)	1.98	0.10	9.36	0.33			0.065	0.004	56.75	<0.01
(4.8)	1.68	0.10	6.50	0.28			0.065	0.004	83.03	<0.01
(4.9)	1.90	0.11	2.55	0.82	0.46	0.09	0.067	0.004	60.24	<0.01
AP-0'										
(4.7)	2.41	0.15	9.14	0.45			0.061	0.004	26.35	3.44
(4.8)	2.16	0.15	5.29	0.33			0.062	0.004	27.29	2.65
(4.9)	2.27	0.16	3.66	1.14	0.22	0.15	0.061	0.004	25.13	3.33
	arc min		ms		ms·deg/s					%

There is no significant difference between the fits for the fully separable and for the fully inseparable blur model. This might indicate that the separable and inseparable components have approximately the same weight in blurring of moving objects. However, by comparing Eq. (4.7) and (4.8), it is easy to see that the similarity between the models could be the simplest explaining factor for the small difference in the goodness of fit between the models. The difference between the model predictions is the smaller the larger the reference blur. With a sharp edge, the difference is at its largest, and even then it is small. This is illustrated in Fig. 4.12, which shows the threshold data for zero arc min and the model fits for observer RO. As a consequence, the errors of the data should be very small before one could make reliable predictions of the degree of separability of blur.

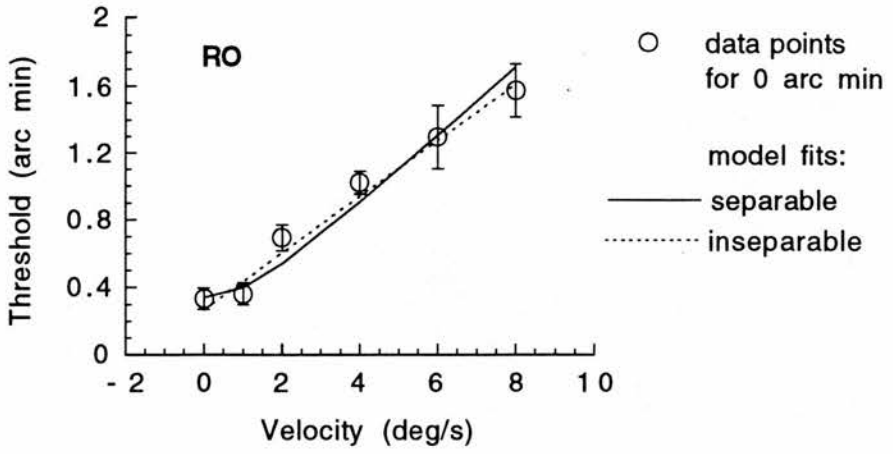


Fig. 4.12. An illustration of the small difference between the fits of the fully separable and inseparable blur models. Observer RO's blur discrimination thresholds for a reference blur of zero arc min are presented together with the model fits.

It is a mathematical certainty that, having one more parameter, the fully inseparable blur model with a quadratic velocity dependence will provide a fit with a smaller χ^2-value than the model with a linear velocity dependence. The difference in the present data, however, is minimal and has no statistical significance. In fact the Q-values for observers RO show a better fit in the linear than in the quadratic case. For observer RO, the value for the second order parameter g is less than half of its standard error, showing also at a glance that g does not significantly differ from zero. The same is true for observers MS and AP (when the data for the reference blur of zero arc min is excluded), even though not equally self-evidently. Thus, the velocity dependence of the effective blur is linear, and the model with a quadratic velocity dependence can be discarded.

The value of the parameter s provides the space constant of the effective static spatial filter. This filter corresponds to the compound effect of all the optical and neural static spatial blurring factors. For observers MS and RO, s is in the range of 0.5 to 0.7 arc min. The same filter for observer AP is considerably larger, having a space constant of about 2 arc min.

Assuming camera-like motion blur, the parameter f is the standard deviation of the temporal weighting function, or the temporal

impulse response of the system. This is why in Table 4.1, the parameter f is expressed in milliseconds. For observers RO and MS, f is in the range of 4 to 6 ms. For observer AP, f varies from 5 to 9 ms. The approximate mean for all observers is about 6 ms.

If our visual system uses special spatial filters in the perception of moving objects, then the parameter f simply indicates how the size of the spatial filter depends on the velocity. In this case, milliseconds are not the best temporal units to show the strength of the dependence. When expressed in units of reference blur and velocity, the value of parameter f varies between 0.23 and 0.56 arc min·s/deg. If the size of the effective static spatial filter is about 0.6 arc min and f is about 0.3 arc min·s/deg, as for observers MS and RO, the effective static spatial blur is larger than the equivalent spatial blur produced by motion (i.e. fv) up to a velocity of about 2 deg/s. For observer AP, the corresponding limit goes to about 4 deg/s. Thus, there is a limiting velocity of 2 to 4 deg/s above which the effects of the velocity-dependent part of blurring dominate. This holds both for the camera-like motion blur and for the case where the size of the filter is assumed to increase with velocity.

The value of the parameter k provides the Weber fraction for the comparison of internal representations of blur. For each observer, the value of k is almost the same for every model. This results from the fact that in every model, it is actually a similar multiplier of the term that contains all the blurring factors. For observers MS and RO, the value of k is about 0.15. Observer AP has a k-value of 0.06, indicating a much better accuracy in blur comparison than the other observers.

The fitting also provided an interesting finding concerning the effects of reference blur. The original experiment carried out by observer AP included not only reference blurs of 0, 1, 2 and 4 arc min, but also 8 arc min. When fitting this data, it was found that the models fit badly to the results for 8 arc min reference blur. To clarify the distinction between 8 arc min and other reference blurs, the thresholds for 8 arc min were compared to the values predicted by the models with parameters obtained from the results for the smaller reference blurs. They were found to be considerably higher than the models predicted. This is illustrated in Fig. 4.13, which also demonstrates the small difference in the predictions of the separable and inseparable blur models for larger reference blurs. The explanation for the inconsistency of 8 arc min with

other reference blurs was not found at that moment, and the measurements for the reference blur of 8 arc min were excluded from the main experiment. The question, however, was revisited later when observer AP repeated the main experiment, and a plausible explanation was found. It will be described in Section 4.10.

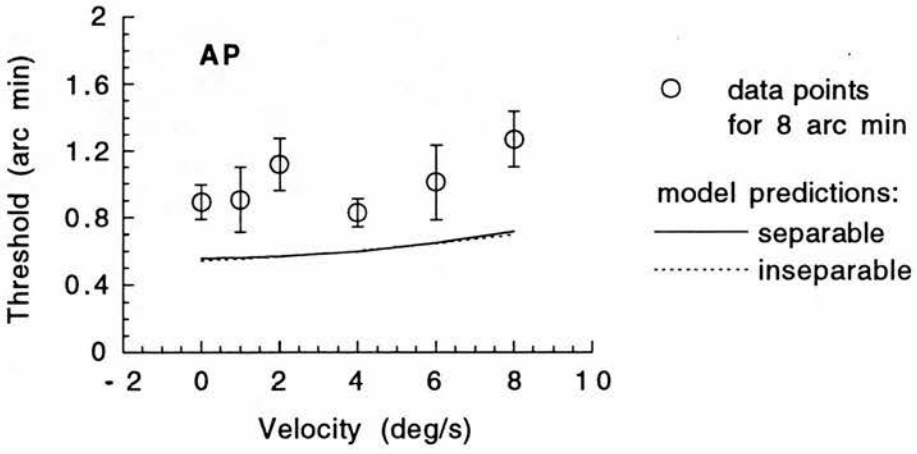


Fig. 4.13. Blur discrimination thresholds as a function of velocity at a reference blur of 8 arc min compared to the values predicted by the models with parameters obtained from the fit to the data for the smaller reference blurs. The thresholds are much higher than the models predict.

4.5.6. Discussion of the results

Effective static spatial filter

The best fit estimates for the space constant of the effective static spatial filter are in the range of 0.5 to 0.7 arc min for observers with normal vision (RO, MS). Levi and Klein (1990) studied the effects of Gaussian blur on 2-line resolution and found that when the stimulus blur exceeds a certain point, thresholds are degraded. They defined this transition point as the equivalent intrinsic blur. According to their measurements, the equivalent intrinsic blur in the fovea is approximately 0.4 - 0.7 arc min. This value is almost equivalent to our estimate, which is exactly what one would expect since the equivalent intrinsic blur is the same as the effective static spatial filter, the former is just expressed in blurring terminology and the latter in filtering terminology. If we assume that our effective static spatial filter

has a form of a Laplacian of a Gaussian, then the estimate for the diameter of the excitatory centre of the filter is in the range of 1.4 to 2.0 arc min. This is consistent with Marr, Poggio and Hildreth (1980), who suggested, from the data on two-point acuity, that the smallest receptive field for a foveal mechanism in human vision must have a central diameter of between 1 and 2 arc min. Our effective static spatial filter is also about the same size as the smallest filter in the pattern discrimination models of Nielsen et al. (1985) and Watt and Morgan (1985), but only about a half of the smallest filter of Wilson and Gelb (1984).

The estimated spatial filter size for observer AP is about three times larger than that for the other observers. Most of this difference might be explained by the acuity of observer AP which is about half of the normal mean value.

Temporal integration

The estimates for the standard deviation of the temporal impulse response of the system provided by the model fits are in the range from 4 to 9 ms. These are maximum estimates based on the assumption that all the blur produced by motion is camera-like. The temporal impulse responses presented in the literature are typically biphasic, having an excitatory phase followed by a smaller inhibitory phase (Watson, 1982). Standard deviation is a measure of probability distributions, and it is not the best descriptor for a weighting function that contains both positive and negative parts. To compare the results with values in the literature, it is assumed (supported by results from some approximate simulation tests) that the estimates reflect mainly the spread of the dominant excitatory part of the impulse response function.

The temporal extent of the impulse response depends strongly on the adaptation level - the impulse response becomes shorter when the retinal illuminance increases. This makes direct comparisons between different studies difficult. Probably the closest lighting conditions to those used in this study were in the study of McKee and Taylor (1984), where overhead fluorescent lighting provided a background luminance of 50 cd/m^2. McKee and Taylor reported their results as best fit parameters to a model proposed by Watson (1982). This model is basically a difference of two gamma functions, and similar models have been successfully used to fit both psychophysical (Levinson, 1968; Bergen & Wilson, 1985) and

neurophysiological data (Baylor, Hodgkin & Lamb, 1974). Using the model and the given parameter values, the standard deviations for the excitatory phases of the temporal impulse response functions were calculated. For the foveal vision of the two observers of McKee and Taylor, values of 5.8 and 6.0 ms were obtained. The same method resulted in a value of 8.3 ms for the temporal impulse response of Watson (1982) based on the data of Roufs and Blommaert (1981), and 9.5 ms for the result of Bergen and Wilson (1985). Estimated from the graphic plot, the excitatory phase of the one of Kelly's (1971) temporal impulse responses that corresponds to an adaptation level close to that in this study has a standard deviation of about 5.5 ms. Taken together, the estimate of about 6 ms is reasonably consistent with the values obtainable from the literature for similar lighting conditions.

The other possibility is that the dominant source of motion blur is the increase in spatial filter size with velocity. If this were the case, then our estimate for the temporal parameter *f* would be an overestimate for the standard deviation of the temporal impulse response because it would also contain the component from the size increase. However, the comparison to the literature values presented above shows that our estimate is more likely to be a slight underestimate than an overestimate. There are three possible explanations for this. First, one could speculate that the temporal impulse response in our task is briefer than in the tasks studied in the literature. This does not seem very likely. Second, some motion-deblurring mechanism could remove the blur in excess to that predicted from the temporal impulse response, i.e. remove the amount of blur corresponding to the increase in spatial filter size. It is hard to see any rational motive to this explanation either. The third explanation is the simplest and the most likely one. It assumes that motion blur is camera-like and not a result of an increase in spatial filter size with velocity.

If we assume that the temporal impulse response has the form proposed by Watson (1982), and that the exposure or motion integration time mainly corresponds to the width of the excitatory phase, then our estimate for the motion integration time is approximately in the range of 20 to 40 ms. However, there is other evidence that motion detectors summate for about 100 ms (Burr, 1981; Burr, Ross & Morrone, 1986). Thus, our estimate of motion blur corresponds to only a fraction of what would be expected from the

summation time. This difference is hard to explain without assuming some kind of blur-free integration mechanism.

Weber fraction

The best fit value for the Weber fraction for observers RO and MS is about 15%. This value is higher than those normally observed in spatial dilation studies. For example, Campbell et al. (1970) found a Weber fraction of 6% for sine-wave period discrimination. Greenlee et al. (1990) found similar values for drifting sine-wave gratings. The values for spatial interval discrimination have also found to be constant at approximately 5% (Westheimer & McKee, 1977; Watt, 1984; Levi & Klein, 1990). The only other blur discrimination study to compare our results with is the one by Watt and Morgan (1983). For the reference blurs above 4 arc min, their data showed blur difference thresholds in the range of 10-20%, and even after an approximate correction for the small band height they used, their thresholds would be still higher than 6%.

If the results of the modelling are correct, observer AP's accuracy in comparing internal representations of blur is more than twice as good as the accuracy of the other observers. The estimated Weber fraction of 6% for observer AP is comparable to the values from spatial dilation studies. However, one should realise that the low Weber fraction of AP may be influenced by the fact that he has a much larger effective spatial filter size and a longer temporal integration time than the observers with fully normal vision. For instance, one biologically motivated explanation might be that if the observers use the same neural population (or at least same-sized populations) in the blur discrimination task, AP's lower spatial resolution would allow more neurones to be used in the comparison, thus making it more accurate. Another explanation is that the longer temporal integration could improve the signal-to-noise ratio in the internal representation, allowing a smaller difference in cue to be discriminated with the same statistical significance. These principles would also serve in making the overall performances of different observers more equal. Taken together, Weber fraction of about 15% is regarded here representative for the blur discrimination task in normal vision.

The difference in Weber fractions between spatial dilation tasks and blur discrimination tasks suggests that the blur discrimination process may use spatial measurements that are not as accurate as the

measurements, for instance, in spatial interval discrimination. To illustrate the issue, let us, for example, assume (Watt & Morgan, 1984) that the effective spatial filter has the form of a Laplacian of a Gaussian, and that the visual system uses peaks and troughs in the output of this filter as spatial primitives. In the case of a Gaussian-blurred edge, the output is biphasic, containing only one peak and one adjacent valley. Watt and Morgan (1983) showed that the distance between the peak and the valley encodes the extent of edge blur in a way very similar to the performance of human observers. In the case of a normal spatial interval discrimination task, i.e. two thin bars separated by more than the extent of the filter, the same filter produces an output that contains two separate peaks, both with flanking troughs making six spatial primitives altogether. Even with equally blurred edge and bars, the peaks from the bars are much narrower, i.e. they have a much better positional accuracy than the peak or the trough from the edge. Even if the accuracies of the positions of the elementary primitives were the same, having fewer primitives (or samples) in the blur discrimination process would mean that the process would have to require a larger difference in cue, i.e. a higher Weber fraction to produce the same statistical significance. This example explanation was based on the use of spatial primitives, but this principle of differences in positional accuracy in explaining the differences in Weber fractions might also hold for a case with no implicit primitives where the comparison is based on the whole filtered output.

Velocity dependence of the effective blur

The modelling showed that the velocity dependence of the effective blur produced by motion is linear. This is an important finding as it shows that there does not necessarily have to be any other type of motion blur than camera-like blur to produce the results. In fact, the strict linearity of the relationship makes it tempting to believe that camera-like blur is at least the dominant component of motion blur. Since the task in our experiment was one-dimensional, in the sense that blur was studied along the direction of motion, the data and the models, however, cannot discriminate between camera-like blur and the case where the motion filter size is a linear function of velocity. This question was addressed in the next experiment.

4.6. Experiment 7: direction of motion perpendicular to blur

The previous experiment showed that the velocity dependence of the effective blur produced by motion is linear. However, since blur was studied only along the direction of motion, it was not possible to discriminate between camera-like blur and the blur produced by an isotropic motion filter whose filter size is a linear function of velocity. The question was investigated by measuring discrimination thresholds for Gaussian-blurred edges with motion trajectories perpendicular to the edge profile. It should be noted that isotropic form is only one assumption for the structure of motion filters. Another assumption is that there exist independent motion filters which are not isotropic, i.e. whose structures or extents are not circularly symmetric. With the non-isotropic assumption, the present experiment could be seen as a test to study the filter size in the dimension perpendicular to motion.

The stimulus arrangements and setup parameter values were the same as in the main experiment. The only exception was that the bands moved perpendicularly to the edge profile, i.e. horizontal bands moved vertically. Observer AP carried out measurements for six different velocities (0, 1, 2, 4, 6, and 8 deg/s) at a reference blur of 2 arc min. In addition, observers AP and MS made some control measurements also at other reference blurs and velocities.

Results and discussion

The results for observer AP at a reference blur of 2 arc min are shown in Fig. 4.14, together with the corresponding results from the main experiment. In contrast to the case where the blur was along the direction of motion, velocity had no discernible effect on the thresholds over the velocity range of 0 to 8 deg/s when the direction of motion was perpendicular to the edge profile. Control measurements of AP and MS confirmed this finding for other reference blurs. The result can be regarded as evidence against any isotropic motion filter since with this kind of filter, the blurring effects should be similar in all directions of motion. One can, however, present another possible explanation why perpendicular motion did not affect the thresholds. The band used was exactly the same as in the main experiment, being 20 arc min in height. With a velocity of 8 deg/s, it takes more than 40 ms from the band to sweep over a horizontal line on the retina, i.e. there is a horizontal stripe on

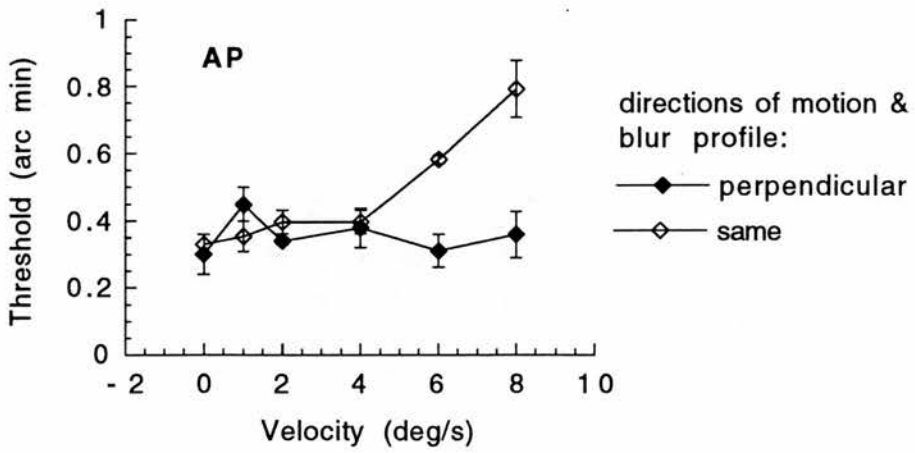


Fig 4.14. Blur discrimination thresholds for observer AP at a reference blur of 2 arc min are shown as a function of velocity for two cases: the direction of motion either perpendicular or the same as that of the edge profile. Velocity has no effect on the thresholds over the examined velocity range in the perpendicular case.

the retina that keeps the same luminance profile for more than 40 ms, and even longer with lower velocities. If the blur-producing motion integration time is less than 40 ms, the visual system could use the same filters as for stationary objects, thus also producing the same amount of blur.

4.7. Masking the information at the ends of trajectories

With experiments of moving targets, one might argue that subjects can use information from stationary glimpses of start and end stations or frames of the trajectory, and that the results do not thus fully represent the motion system. There are several ways to mask the start and end information. In the following experiments, some of these ways are tested, and the effects of masking in motion experiments are discussed.

4.7.1. Experiment 8: zero contrast at the ends of the trajectory

In this experiment, the information about the shape of the edge was removed at the ends of the motion trajectory by making the contrast of the edge zero at those frames. In detail, the contrast was zero during the three first frames, after which it increased in three frames to its setup value. To avoid too abrupt changes, the increase had a form of an integrated Gaussian. At the end of the trajectory, the contrast went through the same changes, but in reversed order. The total number of stations during the 150 ms presentation time in this experiment was sixty, and thus the contrast was practically zero during about 10 ms both at the start and at the end of the trajectory. Other experimental details were identical to those in the main experiment. Two observers (MS and AP) participated in the experiment. Blur discrimination thresholds were measured for six different velocities (0, 1, 2, 4, 6, and 8 deg/s) at a reference blur of 1 arc min.

Results and discussion

The results for both observers are shown in Fig. 4.15. For comparison, the figure also presents the corresponding data from the main experiment. For observer AP, this data is not from the original main experiment, but from the repeated one (described in Section 4.9), which was selected because of its temporal proximity to the present experiment. The same data is also used for comparison in the other experiments of this section.

For both observers, thresholds for the edges having a tapered temporal contrast profile are higher than for those without tapering at and above a velocity of 2 deg/s. For AP, they are higher also at lower velocities, while for MS, thresholds for the tapered and

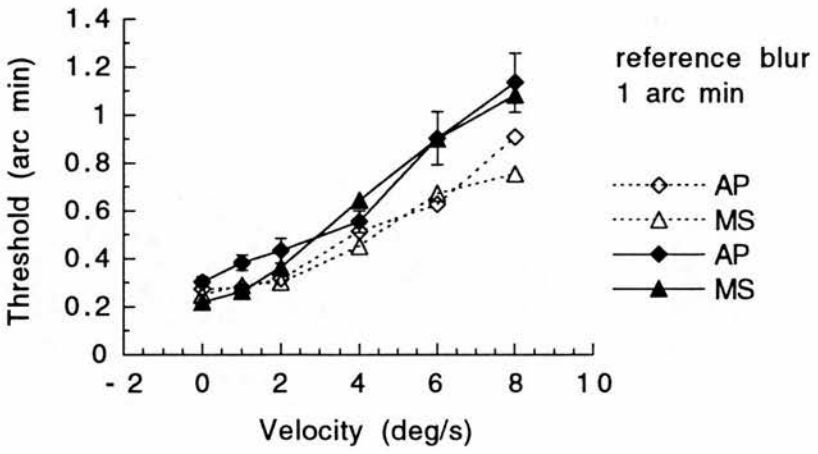


Fig. 4.15. Blur discrimination thresholds for observers AP and MS at a reference blur of 1 arc min are plotted as a function of velocity for edges having a tapered contrast (solid lines, filled symbols) at the ends of the trajectory and for edges without tapering (dashed lines, open symbols). Errors bars for observer AP in the case of tapered contrast are shown as a representative example.

non-tapered case are approximately equal below a velocity of 2 deg/s. The relative difference between the cases increases approximately linearly with velocity.

A simplified interpretation of results would be that since the temporal contrast-profile tapering increases the thresholds, subjects are using information from the stationary start and end frames of the presentation when there is no tapering. There are, however, other possible explanations for the increase in thresholds. Tapering reduces the effective exposure duration, and the shortening of exposure duration is known to lower the performance in many psychophysical tasks, for instance in vernier acuity of moving targets (Morgan, Watt & McKee, 1983), length, orientation, curvature, and stereoscopic depth discrimination (Watt, 1987), and velocity and direction discrimination (De Bruyn & Orban, 1988). Another explanation is that the tapering not only removes the spatial information at the ends of the trajectory, but also changes the velocity information. Bringing contrast to zero can be seen as transferring luminous "mass" from the high-luminance end of the band to the low-luminance end. This will increase the perceived velocity at the end of trajectory in the case where the low-luminance end is

leading in motion. The impression of the observers was indeed a sudden acceleration of the edge at the end of trajectory, which also appeared to produce a large amount of blur. This appearance is consistent with the idea that the perception of form and motion cannot be separated. If this idea mirrors reality, it is impossible to mask the information about the spatial form of the object without affecting the motion information, and because of that, without distorting the perceived spatial form information itself.

There is evidence from experiments with moving vernier targets (Westheimer & McKee, 1977; Morgan, Watt & McKee, 1983) that the form of a moving target is not determined just by brief glimpses of a limited portion of the trajectory. On the basis of this evidence and the above discussion, it is concluded here that the reason for the increased thresholds produced by contrast tapering is not the masking of the stationary start and end frames, but the reduction and distortion of the overall information about the form and motion of the edge.

4.7.2. Experiment 9: Gaussian luminance profile along the trajectory

The relative weight of the information about the shape of a target that comes from the start and end frames of its motion trajectory can be reduced by lowering the average intensity of the target selectively at these parts of its motion path. In this experiment, the average luminance of each band was adjusted so that it was a Gaussian function of time, and thus also of spatial position. The standard deviation of the Gaussian profile was 30 ms, i.e. one fifth of the exposure duration. The average luminance of the first and the last frame corresponded to the value of a normalised Gaussian function at a distance of 2.5 standard deviations from the mean, i.e. to about 4% of the peak luminance. The average luminance of the band increased from this small value at the first frame to the peak luminance in the middle of the trajectory, and then decreased back to the same small value. The peak average luminance was adjusted so that the total energy of each band was the same as in the main experiment. Other experimental details were identical to those in the main experiment. Blur discrimination thresholds were measured for six different velocities (0, 1, 2, 4, 6, and 8 deg/s) at a reference blur of 1 arc min by two observers (MS and AP).

Results and discussion

The results for both observers are shown in Fig. 4.16. Drawn also in the figure are the corresponding thresholds from the main experiment for the same observers (from the repeated main experiment for AP). When the average luminance of a moving edge follows a Gaussian function over time, blur discrimination thresholds are higher than when the edge has its average luminance constant along the trajectory. For observer AP, the Gaussian adjusted average luminance also resulted in a higher threshold for a stationary edge. For both observers, the increase in threshold as a function of velocity produced by the luminance adjustment is much stronger than the linear increase produced by the contrast tapering in the previous experiment.

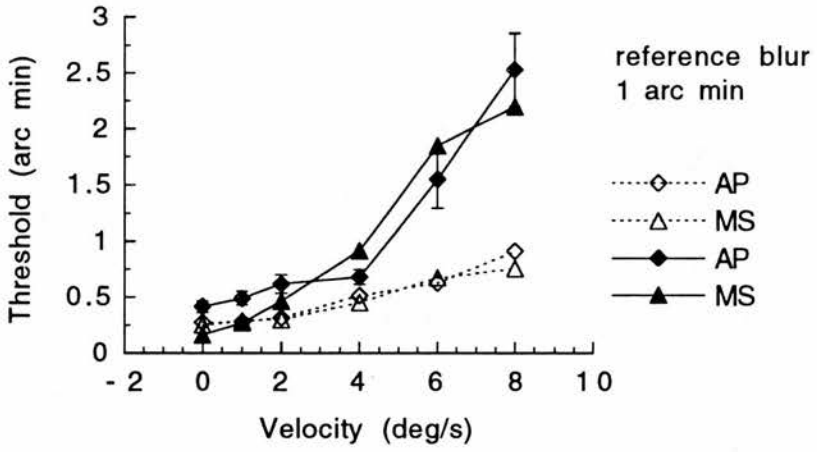


Fig. 4.16. Blur discrimination thresholds for observers AP and MS at a reference blur of 1 arc min for edges having their average luminances adjusted along the trajectory according to a Gaussian function (solid lines, filled symbols). The corresponding thresholds from the main experiment, i.e. with the average luminances constant along the trajectory, are plotted for comparison (dashed lines, open symbols). The luminance adjustment produces a progressive increase in thresholds with velocity. Errors bars for observer AP in the case of Gaussian luminance profile are shown as a representative example.

The possible explanations for the increase in thresholds produced by Gaussian temporal luminance adjustment are in general the same as those suggested for the contrast tapering. First, luminance adjustment reduces the effective exposure duration. Second, it distorts both the velocity and the spatial form information about the edge by adding luminance from "nowhere" in the first half of the

trajectory and subtracting it in the second half. The observers perceive simultaneously a flash and a movement, without a proper impression of a rigid object. It is difficult to say whether these suggestions could be used to explain why the thresholds increase faster than linearly with velocity, or if other explanations are needed.

4.7.2. Experiment 10: moving edge viewed through a window

If we look at an object moving behind a fence, our visual system can synthesize the form of the object even if the whole of it is never simultaneously in view. In this experiment, the information about the shape of a moving edge coming from the "stationary" start and end frames was removed by letting the edge move behind an opaque "wall" having only one slit or window though which the object was viewed. In this way, there was no start or end frame where the edge could have been seen as a whole, instead the edge came into view from behind the border of the window and disappeared at the other border. The edge could only be seen as a whole while it was moving.

The width of the viewing window was adjusted so that it took 150 ms of the edge to cross it, and thus the width was dependent on the velocity of the edge: the higher the velocity, the wider the window. The crossing time was defined as the time interval between the moments when the midpoint of the edge passed the borders of the window. Thus, the time during which at least a part of the edge was in view was longer than 150 ms, but correspondingly the time when the whole edge was simultaneously in view was shorter than 150 ms. In the presentation of each edge, there was a period of 250 ms when only the low luminance plateaux part of the edge was in sight before the midpoint of the edge reached the window border, and after the edge had crossed the window, there was a similar 250 ms time period when only the high luminance plateaux part was in sight. This presentation imitated a physical situation where an edge moved along a very long band of which only a fraction could be seen. This was different from the presentation of the main experiment where the band moved with the edge fixed to it. However, the total information about the shape of the edge integrated over the presentation of the edge was the same in both experiments.

Other experimental details than those described above were identical to those in the main experiment. Two observers (MS and AP) participated in the experiment. Blur discrimination thresholds

were measured for four different velocities (2, 4, 6, and 8 deg/s) at a reference blur of 1 arc min.

Results and discussion

It was found that viewing the edges through a narrow window - for example at a velocity of 2 deg/s, the width of the window was only 18 arc min - makes the discrimination task very difficult, producing a marked increase in thresholds. In fact, observer AP also tried to measure the thresholds for a velocity of 1 deg/s, leading to a window width of 9 arc min that is only about 50% wider than the whole extent of an edge with a standard deviation of 1 arc min. No clear impression of the spatial form of the moving edge was perceived, which was why this velocity was excluded from the experiment. However, the percepts of form were not very clear even at higher velocities.

Fig. 4.17 shows the thresholds at a reference blur of 1 arc min for moving edges viewed through a window for both observers, together with the thresholds for the same reference blur from the main experiment. At a velocity of 2 deg/s, the thresholds for edges viewed through a window are three times as high as those in the main experiment for both observers. The ratio drops to about two at 6 deg/s, but then increases again above two and a half at 8 deg/s.

The experiences of the observers in this experiment show that the capacity of our visual system to synthesize the form of an object even if all parts of it are never simultaneously in view largely fails in this specific situation. Several suggestions are presented here to explain the increase in thresholds. First, it is suggested that the integration of form works more efficiently when the whole object is simultaneously in view. The borders of the window are luminous edges that distort the form of an edge when it passes the borders. Since the amount of time of seeing the moving edge as a whole is shorter at lower velocities, the detrimental effects should also be stronger at low velocities. Second, the form integration system might even summate the edges produced by the window borders on the moving edge. This hypothetical process would be the more harmful the closer the border edges are to the moving edge. Both in spatial and velocity domains, they are closer to the moving edge at low velocities. There is some indirect evidence for this suggestion from a study of Bowne, McKee and Glaser (1989). They found that human speed discrimination can be degraded by additional stimuli

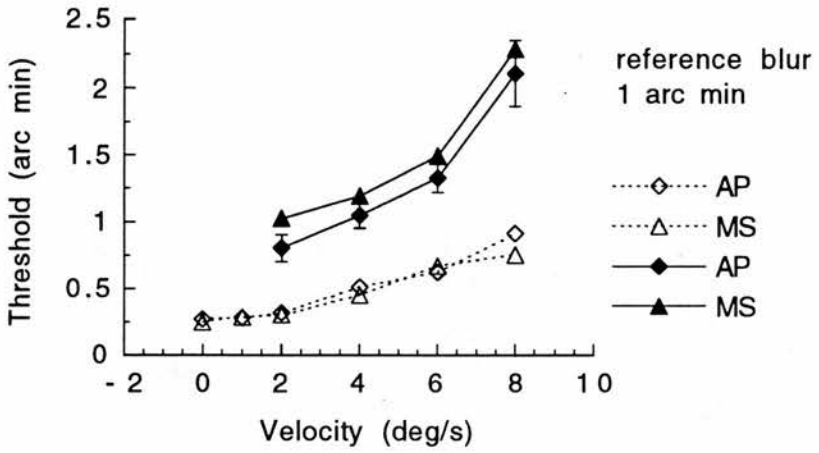


Fig. 4.17. Blur discrimination thresholds are plotted as a function of velocity for moving edges viewed through a window for observers AP and MS (solid lines, filled symbols). The edges had a reference blur of 1 arc min. At each velocity, the width of the window was such that it took 150 ms at that velocity to cross it. For comparison, the thresholds for a reference blur of 1 arc min from the main experiment where the edges moved fixed to a constant-sized band are also shown (dashed lines, open symbols). For all the velocities tested, the thresholds for edges viewed through a window are at least twice as high as those in the main experiment. Error bars for observer AP in the case of edges viewed through a window are shown as a representative example.

in close spatial and temporal proximity to designated test target. If we assume that the analysis of both the spatial form and the velocity of a moving object are inseparable parts of our motion system, then an interference in the analysis of velocity would also produce an interference in the analysis of form.

The two suggestions presented above can be used to explain the marked increase in thresholds at low velocities. They also explain the decrease in the ratio between the thresholds of this and the main experiment over the velocity range of 2 to 6 deg/s. However, they don't predict the fast increase between 6 and 8 deg/s. For this, a third explanation is needed that has its strongest effects at high velocities.

4.8. Experiments on velocity information

The models of motion deblurring (Burr, Ross & Morrone, 1986; Anderson & van Essen, 1987) rest on the principle that positional acuity is restored by taking into account the temporal delay at which different photoreceptors have been stimulated. To accomplish this accurately, these models assume that the motion of the object is uniform and rectilinear. A more general principle for motion deblurring would be complete deconvolution which has no special velocity requirements except that velocity has to be known at each moment of the temporal integration period. Deconvolution would not increase the total spatial information obtained from a moving target - it would only use velocity information to transform spatial information from a blurred form to a non-blurred form. It is difficult to infer which form, a blurred form combined with velocity information or a non-blurred form, the visual system would prefer in subsequent analysis. In any case, deconvolution would be very hard to accomplish with neural hardware, and if motion-deblurring exists, it is more likely to be realized by a method that relies on uniform and rectilinear motion.

The first two experiments of this section study how a non-uniform motion will affect blur discrimination thresholds. The third experiment is different in its purpose: it studies how the velocity information coming from the ends of the moving bands affects the thresholds.

4.8.1. Experiment 11: accelerating target

The human visual system is believed to integrate or average velocity information over time. This idea has been inspired by the finding that humans are unaware of modest target accelerations and decelerations (Gottsdanker, 1956; Schmerler, 1976; Morgan, 1980; Snowden & Braddick, 1991). This experiment was designed to study how acceleration or deceleration of a moving edge affects the integration of its spatial form.

Blur discrimination thresholds were measured for moving edges with a constant acceleration or deceleration. The mean velocity averaged across the whole trajectory was 4 deg/s. Five different velocities at the start of the trajectory were used: 0, 2, 4, 6, and 8 deg/s, and thus the velocities at the end of the trajectory were

8, 6, 4, 2, and 0 deg/s, respectively. Since the exposure duration was 150 ms, the corresponding accelerations were 53.3, 26.7, 0, -26.7 and -53.3 deg/s^2. Other experimental details were the same as those in the main experiment. Observers AP and RO measured blur discrimination for a reference blur of 2 arc min. In addition, observer AP measured thresholds for a reference blur of zero arc min.

Results and discussion

The results for accelerating and decelerating edges with a reference blur of 2 arc min are shown in Fig. 4.17 for both observers. With a motion-tuned mechanism, one might expect that acceleration would not allow the mechanism to keep in step with the motion of the edge, thus smearing its spatial representation. In Fig. 4.17, this would mean an increase in threshold on both sides of 4 deg/s which represents the case of zero acceleration. This increase is not evident. Instead, blur discrimination thresholds decrease with increasing start velocity, and thus with decreasing end velocity, reaching their minimum at an end velocity of zero, i.e. when the edge is stationary at the end of the trajectory. This

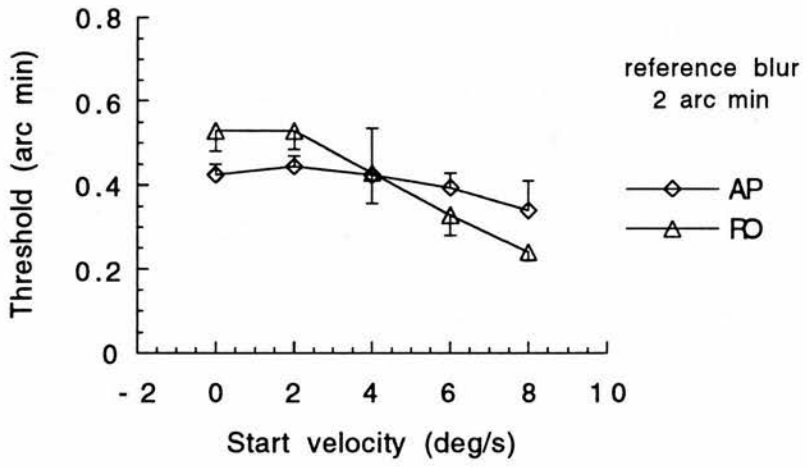


Fig. 4.17. Blur discrimination thresholds for accelerating Gaussian blurred edges are plotted as a function of velocity at the start of the trajectory for two observers (AP and RO). The mean velocity along the trajectory was 4 deg/s, and the reference blur was 2 arc min. Error bars represent one standard error. For both observers, thresholds decrease with increasing start velocity, and thus with decreasing end velocity, indicating such an averaging of the spatial information where more weight is given on the second than on the first half of the trajectory.

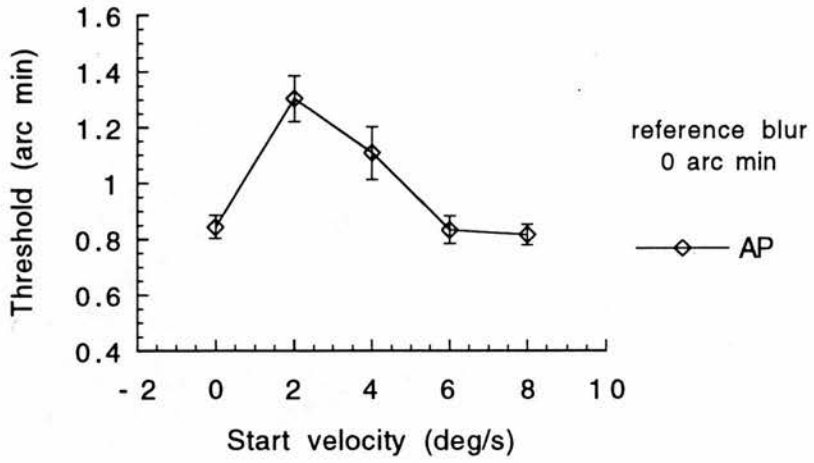


Fig. 4.18. Blur discrimination thresholds for accelerating sharp edges (reference blur 0 arc min) as a function of velocity at the start of the trajectory for observer AP. The mean velocity along the trajectory was 4 deg/s. Error bars represent ± one standard error. When the edge is stationary at the beginning of its trajectory (start velocity 0 deg/s), the threshold is much smaller than that which would be predicted on the basis of a simple integration mainly over the second part of the trajectory.

decrease indicates that the information about the shape of the edge is integrated over a time period that is less than the exposure time of 150 ms, and that the presentation of the edge that is integrated last is used for blur discrimination. If the integration time were equal to the exposure time, then there would be no changes in threshold as a function of start velocity.

Since the blur discrimination thresholds for a reference blur of 2 arc min increase approximately linearly with velocity (see the main experiment), the decrease in threshold in this experiment should be a linear function of start velocity, and this is indeed found in the range of 2 to 8 deg/s. The rate of the linear decrease is faster for observer RO than for AP, indicating a shorter integration time. For both observers, there is a deviation from this linearity at a start velocity of zero deg/s, i.e. when the edge is stationary at the beginning of its trajectory, the threshold is smaller than that which would be predicted on the basis of the rest of the data. This deviation is much more distinct in the data for a reference blur of zero arc min for observer AP which is shown in Fig. 4.18. The thresholds do not differ in the cases where the edge was stationary at the start or at the end of its trajectory. The

explanation suggested here is based on the sensations of observer AP during the experiment. When the edge was stationary at the beginning of its trajectory, AP had the impression that he could separately perceive the form of the stationary edge, without much interference from the consecutive perception of the moving edge. No such impression was perceived at start velocities above zero. This finding evokes the idea that there might be separate representations for moving and stationary targets in our visual system. Without the notion of representations, this idea was originally proposed by Tolhurst (1973) in the form of movement-dependent and movement-independent channels. A perhaps not so distant analogy to this finding is that of Watson and Robson (1981) who observed that two temporal frequencies of flickering gratings can be perfectly discriminated only if one is higher than about 8 Hz and the other less than 2 Hz.

4.8.2. Experiment 12: jittering velocity

Accelerating motion is non-uniform in terms of physics, but if the acceleration is constant as in the previous experiment, velocity either increases or decreases continuously and smoothly with a constant rate. It was found that blur discrimination is rather insensitive to this kind of changes in velocity. The purpose of the next experiment was to study how velocity jitter, i.e. random and sudden changes in the momentary velocity of the edge, would affect blur discrimination thresholds.

Jittering of the velocity was accomplished by selecting the spatial position of the target at each station of the trajectory randomly from a uniform probability distribution centred at the position predicted by the mean velocity. The interstation distance calculated from the mean velocity was selected as a unit for the velocity jitter. For example, if the velocity jitter had a value of three, the spatial position of each station could have any position within three interstation distances from the position predicted by the mean velocity. In this case, the maximum possible momentary forward velocity could be seven times the mean velocity. This would occur if the random positional difference was -3 units at one station and 3 units at the following station, leading to a forward jump of 7 units. Similarly, the maximum possible momentary backward velocity could be five times the mean velocity.

The mean velocity used in the experiment was 4 deg/s. Since there were 30 stations during the exposure time of 150 ms, one unit of the velocity jitter was equivalent to 1.2 arc min. Blur discrimination thresholds were measured at a reference blur of 2 arc min for five different values of velocity jitter. These values were 0, 0.5, 1, 2, and 3. The observer was the author (AP).

Results and discussion

The results (Fig. 4.19) show that blur discrimination thresholds increase with velocity jitter, but only to a modest degree. For higher jitter values, the increase may be even less than that predicted on the basis of the additional stimulus variance produced by the velocity jitter. The initial predicted threshold at zero jitter in Fig. 4.19 is calculated using the best fit parameters of the fully inseparable blur model from Eq. (4.8) to the data of observer AP in the repeated main experiment, but since the measured value at zero jitter is almost the same, it could have been used equally well. For the rest of the prediction curve, it is assumed that jitter adds to the sum of variances of the internal representation of the edge described in Eq. (4.1), and the threshold

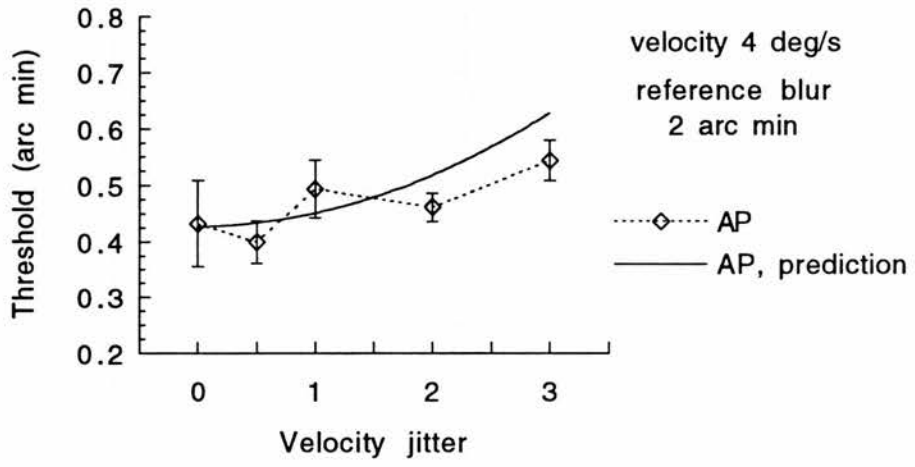


Fig. 4.19. Blur discrimination thresholds for moving, Gaussian blurred edges are plotted as a function of velocity jitter (for definition, see text) for observer AP (open diamonds, dashed line). The edges had a reference blur of 2 arc min. For comparison, a prediction based on the additional stimulus variance produced by the velocity jitter is presented (solid line). The observed thresholds follow the prediction, except that at higher jitter values, they seem to be smaller than predicted.

is then calculated according to the Eq. (4.8). However, since the variance from jitter is separable from the other variances in the internal presentation, any of the Eqs. (4.7)-(4.9) would have produced the same curve. The variance of the stimulus jitters was calculated using the knowledge that the variance of a uniform distribution of width $\pm w$ is $w^2/3$.

The fact that thresholds are fairly insensitive to velocity jitter allows us to make some predictions of the operation of possible motion-blurring mechanisms. Let us for instance assume that the mechanism behind motion deblurring is the linear shifter circuit principle of Anderson and van Essen (1987). The velocity feedback signal in the shift control is an essential part of that model. Since there is no biologically plausible way for the circuit to have a feedback loop fast enough to keep in step with the jittering velocity, the only way to produce the results in Fig. 4.19 is to average the velocity signal considerably before using it in the shift control. This averaging cannot be done without producing delay in the feedback loop. This would mean that for objects changing their mean velocity with time, the shift would always lag behind the correct value, thus producing blur.

An interesting finding is that the thresholds seem to be lower than the predicted values when the velocity jitter is at or above two units. A part of the possible difference can be explained on the basis of the estimate for the motion integration time in the main experiment. This is about 25 ms. There can be approximately only five stations of the motion trajectory inside this time. The variance resulting from five samples is less than the variance resulting from an infinite number of samples that was used in the prediction as a close approximation of about 20 to 30 samples corresponding to the total summation time. This explanation would also require that the possible second-phase integration of the 25 ms packets should be translation-invariant, i.e. it should happen without blur. One should note that for the receptive fields of Burr et al. (1986) having their overall temporal extent over 100 ms, the minimum possible thresholds would be approximately those in the prediction in Fig. 4.19. However, one should also note that since there are results only for one observer in this experiment, the explanation presented above for the difference in thresholds should be taken only as tentative.

4.8.3. Experiment 13: moving edge in a stationary band

There is accumulating evidence that the human visual motion system analyses form and motion of moving objects simultaneously, in a single process where changes in velocity can affect spatial form information and vice versa. In most of the previous experiments, the Gaussian blurred edge has moved together with the band, i.e. they have formed a single moving object. In this object, the velocity information from the ends of the band is more distinct than that from the edge. One might argue that if the velocity information from the ends of the band is removed, blur discrimination thresholds for the edge will increase as a consequence of the fact that the reduced velocity information will lead to a less accurate analysis of the spatial form. This hypothesis was tested in this experiment.

The velocity information from the ends of the band was removed by keeping the band stationary on the screen during each presentation. The length of the band was 200 arc min, i.e. it was 60 arc min longer than in the main experiment. This was because the edge had to have space to move inside the band. The minimum possible distance from the midpoint of the edge to the end of the band was 64 arc min, compared to the fixed distance of 70 arc min in the main experiment. All other experimental details were the same as those in the main experiment. Two observers (MS and AP) participated in the experiment. Blur discrimination thresholds were measured for six different velocities (0, 1, 2, 4, 6, and 8 deg/s) at a reference blur of 1 arc min.

Results and discussion

The results for both observers are shown in Fig. 4.20, which also presents the corresponding data from the main experiment (from the repeated one for AP). At and above a velocity of 2 deg/s, thresholds for edges moving along a stationary band are higher than for those moving with a band, and they increase approximately linearly with velocity. The distinction between velocities below and above of about 1.5 deg/s is interesting. It may reflect a separation in the spatial analysis of moving and stationary (or slowly moving) targets.

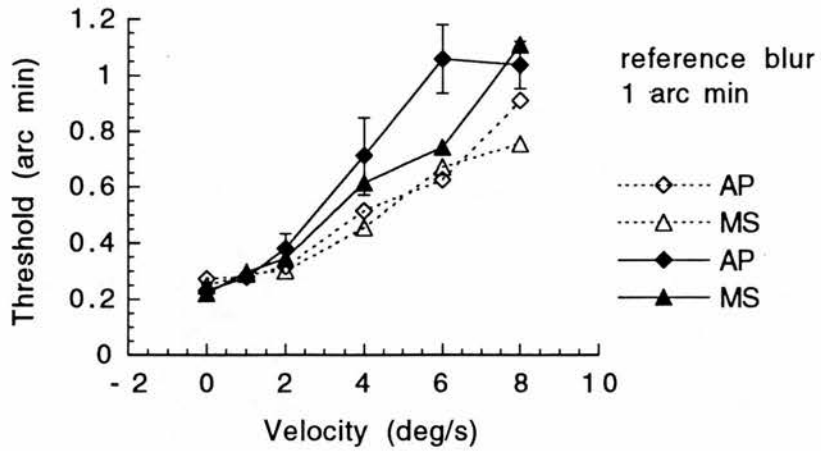


Fig. 4.20. Blur discrimination thresholds for observers AP and MS at a reference blur of 1 arc min are plotted as a function of velocity for edges moving along a stationary band (solid lines, filled symbols) and for edges moving with a band (dashed lines, open symbols). Errors bars for observer AP in the case of stationary band are shown as a representative example.

The results are consistent with the hypothesis that the removal of the velocity information from the ends of the band will increase blur discrimination thresholds. It cannot be specified whether this increase is a consequence of a less accurate spatial analysis due to the reduced velocity information, or if the stationary ends just interfere more with the spatial analysis of the edge than the ends moving with it. In any case, it is suggested that the results provide some evidence for the inseparable analysis of form and motion of moving targets.

4.9. Experiment 14: main experiment repeated

It was noticed during the experiments with masked information at the ends of trajectories and during the experiments on velocity information that thresholds for observer AP might have changed from the main experiment. For this reason, observer AP repeated the experiment. Six months had gone since he had carried out the main experiment for the first time. All the experimental details of the experiment were the same as those in the main experiment.

Results

Fig. 4.21 shows blur discrimination thresholds as a function of velocity for four different reference blurs for observer AP. The main features are the same as in the original main experiment. For each reference blur the thresholds increase approximately linearly with velocity, and the rate of this increase is inversely related to the reference blur. At each velocity, the thresholds for reference blurs of 1 and 2 arc min are smaller than the corresponding ones in the main experiment (Fig. 4.10). For a reference blur of 1 arc min, the average decrease is about 0.05 arc min. For 2 arc min, the decrease is slightly smaller. For a reference blur of zero arc min, the change is in the same direction, but it manifests only at velocities of 0, 6, and 8 deg/s. These results show that observer AP's performance in blur discrimination of small reference blurs has improved during the time period between the experiments.

The most remarkable finding is, however, for a reference blur of 4 arc min. As Fig. 4.22 shows, the thresholds for this reference blur are on the average about 0.15 higher in this experiment than they were in the original main experiment. The threshold also has a stronger velocity dependence than in the main experiment. Thus, it seems as the visual system is paying a price for the improved performance at small reference blurs: the performance at higher blurs declines. To understand what changes in the blur discrimination system could have caused the changes in thresholds, the models described in Eqs. (4.7), (4.8) and (4.9) in Section 4.5.2. were fitted to the data of this experiment.

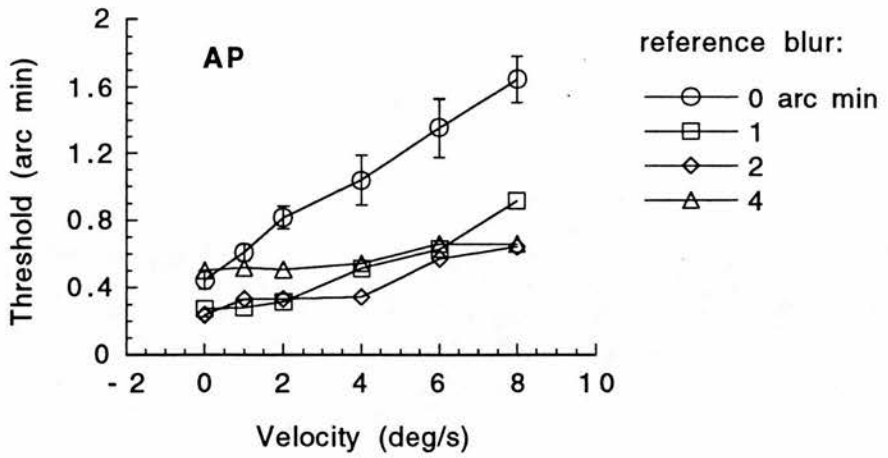


Fig. 4.21. Blur discrimination thresholds in the repeated main experiment as a function of velocity at four different reference blurs for observer AP. Error bars for zero arc min reference blur are presented as an example, they represent ± one standard error. One standard error was typically about 10% of the threshold. The thresholds for reference blurs of 0, 1, and 2 arc min are smaller than the corresponding ones in the main experiment (see Fig. 4.10).

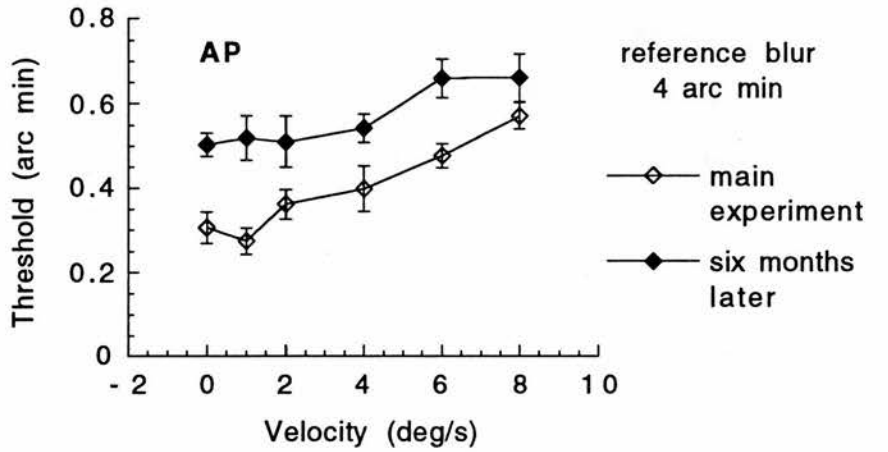


Fig. 4.22. Blur discrimination thresholds for a reference blur of 4 arc min from the main experiment (open diamonds) compared to the thresholds from an exactly identical experiment (filled diamonds) carried out by the same observer (AP) six months later. In contrast to the change at smaller reference blurs, this data shows a decline in performance between the experiments.

Results of model fitting

The same method as in the main experiment, i.e. the Levenberg-Marquart method briefly described in Section 4.5.4, was used to fit the models to the data. The principles for checking the validity of the fitted parameters were also the same as in the main experiment.

Fig 4.23 shows the fit of the fully inseparable blur model (Eq. (4.8)) superimposed on the data. The model fits to the data well, as do also the other models. The Q-values indicating the goodness of fit are presented in Table 4.2 together with other results from the fitting. For comparison, the corresponding results from the original main experiment are also presented in the table. The Q-values for observer AP in this experiment are considerably higher than in the main experiment. This is mainly due to much better fits to the thresholds for a reference blur of zero arc min, though this reference blur is still the one providing the largest sums of squares of deviations in the merit function.

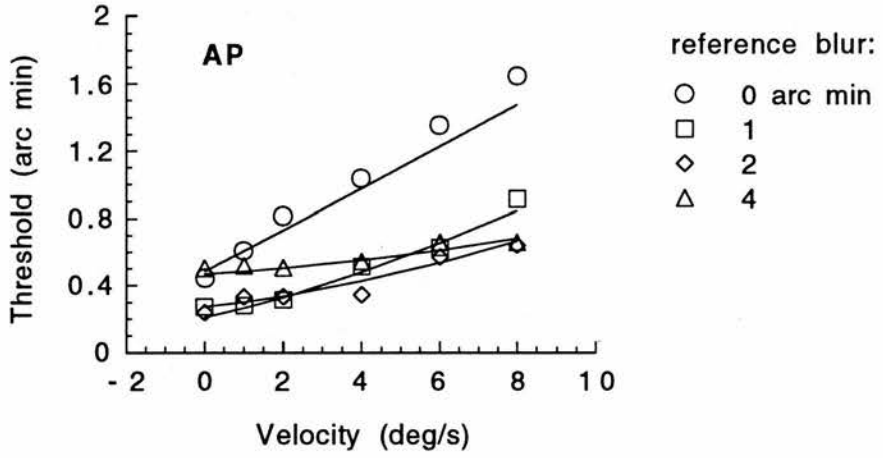


Fig. 4.23. A fully inseparable blur model fit superimposed on the blur discrimination threshold data for observer AP in the repeated main experiment.

The best fit estimates for the space constant of the effective static spatial filter (parameter s) show a decrease from almost 2 arc min down to about 1 arc min. Also the estimates for the standard deviation of the temporal impulse response (parameter f in Eqs. (4.7) and (4.8) when blur is assumed to camera-like) show a decrease. The size of this decrease is about 30%. On the other hand, the estimate for the Weber fraction for the comparison of internal representations of blur (parameter k) has almost doubled

Table 4.2. Comparison of the results from the modelling of blur discrimination thresholds for observer AP in the main experiment with those in the same experiment repeated six months later. Presented are best fit parameters, their standard errors, χ^2- and Q-values for three different models (labelled according to the equation numbers in text: (4.7) fully separable blur, (4.8) fully inseparable blur with a linear velocity dependence, and (4.9) fully inseparable blur with a quadratic velocity dependence).

	s	σs	f	σf	g	σg	k	σk	χ^2	Q
main										
(4.7)	1.98	0.10	9.36	0.33			0.065	0.004	56.75	<0.01
(4.8)	1.68	0.10	6.50	0.28			0.065	0.004	83.03	<0.01
(4.9)	1.90	0.11	2.55	0.82	0.46	0.09	0.067	0.004	60.24	<0.01
rept.										
(4.7)	1.16	0.07	6.13	0.27			0.111	0.004	31.98	5.89
(4.8)	1.00	0.07	4.27	0.27			0.111	0.004	24.05	29.05
(4.9)	1.02	0.08	3.85	0.91	0.05	0.11	0.111	0.004	23.82	25.03
	arc min		ms		ms·deg/s					%

its value from 0.065 to 0.111. The Q-values for the fully inseparable bur models show a better fit in the linear (Eq. (4.8)) than in the quadratic case (Eq. (4.9)), confirming the conclusion made in the main experiment that the velocity dependence of the effective blur is linear. As a whole, the best fit parameter sets for observer AP have changed towards the values obtained for the other two observers in the main experiment.

Discussion

The results of the model fitting give information about the nature of learning in blur discrimination. The change in the Weber fraction shows that the accuracy for the comparison of internal representations of blur declined from the main experiment to the repeated experiment. This is surprising - one might have expected that learning would have improved especially the comparison of blurs. However, the results show that the blur discrimination system has accomplished the improvement for small reference blurs by changing other factors in the process. It is difficult to say whether the increase in the Weber fraction is a consequence of resharing of limited resources, or whether there are some simple principles of decision making statistics behind it. For example, if

we assume that blur discrimination is based on the comparison of peaks and troughs in a Laplacian of a Gaussian filtered signal, then having a smaller spatial filter would lead to a smaller number of neurones representing the primitives. To reach the same statistical significance in decision, the system would have to require a larger difference in cue, and thus a higher Weber fraction.

The decrease in the estimated space constant of the effective static spatial filter both explains and gets its explanation from the finding that observer AP's performance in blur discrimination improved for small reference blurs and declined for a reference blur of 4 arc min. A practical reason for this specific learning is that observer AP did a large number of other blur discrimination experiments between the original main experiment and the repeated experiment using a reference blur of 1 arc min. The reason for the decrease in the estimated standard deviation of the temporal impulse response, and thus also in the elementary temporal integration time may just be that these other experiments were mostly done with moving targets.

From a point of view of the neural implementation of blur discrimination system, the results are particularly interesting. They strongly indicate neural plasticity in the mechanism producing internal representations. The changes reflect either growth of new connections or, more likely, alterations in the strengths of previously existing connections. There is much evidence from neurophysiological studies using receptive field mapping that not only developing brain but also adult visual cortex can exhibit significant reorganisation (Kaas, Krubitzer, Chino, Langston, Polley & Blair, 1990; Gilbert & Wiesel, 1991; Chino, Kaas, Smith III, Langston & Cheng, 1992, among others). In these experiments , the changes have been induced by restricted retinal deafferrentiation, i.e. by producing local lesions in the retina. The reorganization is a learning effect: a lesion produces a zero visual input, and the visual cortex learns to give a zero weight to the connections coming from the lesioned area.

In psychophysics, practice effects are known to be quite pronounced, for example for vernier targets. It is a common routine to practice tasks extensively before starting of data collection. These practice effects are generally assumed to be a consequence of rather high-level learning, for instance of forming and refining of internal memory templates of target configurations. This view,

however, is not shared unanimously. Recently, Poggio, Fahle and Edelman (1992) presented a simple neural network model that learned to solve a vernier task at a hyperacuity level. In any neural network model, learning is equivalent to optimizing the connection strengths or weights between the neurones of the model. The success of the model of Poggio et al. demonstrates that learning in a psychophysical task could happen at a low level through neural plasticity. Learning in their model is fast, but it is proposed here that a much more complicated true biological network can also show low-level learning effects on a much slower time scale. The results of the present experiment are suggested to be one sign of this.

4.10. Experiment 15: effect of reference blur

The blur discrimination models described in Section 4.5.3 (Eqs. (4.7)-(4.8)) predict that for stationary edges with reference blurs larger than the optimum blur, blur discrimination thresholds are directly proportional to the reference blur. This is different from the results of Watt and Morgan (1983b), who found that beyond the optimum level, thresholds rise corresponding approximately to a power law with an exponent of 1.5. Some of the observations in the original and the repeated main experiment led to a likely explanation of why Watt and Morgan found that thresholds rise faster than expected from a simple Weber's Law relation. In the original main experiment, observer AP measured thresholds not only for reference blurs of 0, 1, 2 and 4 arc min, but also for 8 arc min. The models, however, fit badly to the results for 8 arc min reference blur: the thresholds were considerably higher than the models predicted. This has been shown earlier in Fig. 4.13 in Section 4.5.5. Up to the reference blur of 4 arc min, the predictions were good.

To find out what makes the difference between 4 and 8 arc min blurs, the real blur extents were calculated. Since the reference blur space constants are expressed in standard deviations and the practical extension for a Gaussian is about ± 2 standard deviations, the total blur extents for 4 and 8 arc min reference blurs are 16 and 32 arc min, respectively. Both values are well inside the 120 arc min length of our stimulus band, but 32 arc min is far larger

than the 20 arc min height of the band. It was assumed that this might be the reason for the degraded thresholds. For instance, if the total integration area for the blur discrimination system were circular with a diameter of the total blur extent, edges with total blur extents larger than 20 arc min would fill only a part of the area, resulting in an inadequate integration of the blur and thus, raising the thresholds. There is some evidence for this kind of integration from grating visibility experiments (Howell & Hess, 1978; Robson & Graham, 1981) where it has been found that the contrast threshold of sine-wave gratings is influenced by the spatial extent of the grating in both horizontal and vertical directions. The detectability of a grating increases with the number of cycles presented up to a critical number, and this critical width of a patch of vertical grating is the same as the critical height. The results of Levi and Klein (1986) in a spatial discrimination task also provide some support for the assumption. In a 3-line bisection experiment, they found that increasing the number of samples in the direction orthogonal to the discrimination cue reduces the thresholds, up to a limiting number of five samples for foveal vision.

To test the assumption, observer AP measured blur discrimination thresholds for the 8 arc min reference blur with a 40 arc min band height. The doubling of the band height was done by halving the viewing distance to 50 cm - due to technical difficulties the physical height of the band could not be increased. All other parameters and experimental conditions were the same as in the main experiment.

Results and discussion

The results are shown in Fig. 4.24. The thresholds are close to the values predicted by the fully inseparable blur model with parameters obtained from the results for the smaller reference blurs. The consistency of the data was also tested by including the thresholds for the reference blur of 8 arc min in the data and refitting it. The fit was very close to that with 8 arc min excluded - no parameter changed its value by more than 4%. The same was true also for the fully separable blur model.

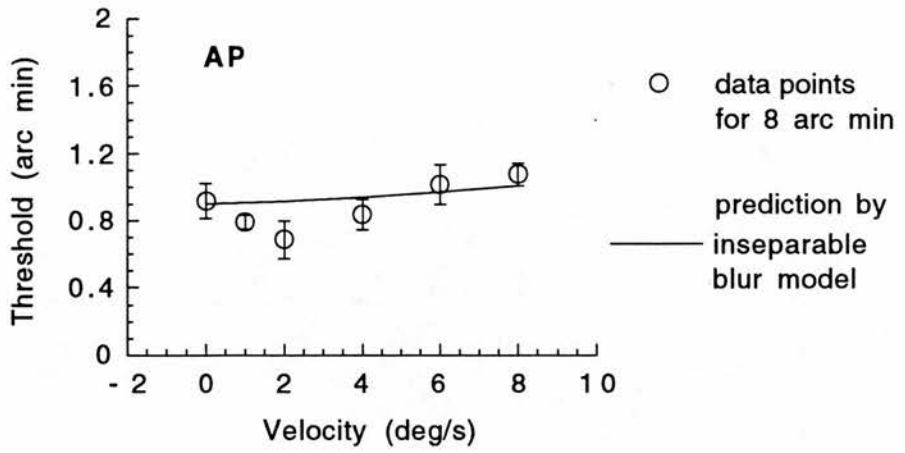


Fig. 4.24. Blur discrimination thresholds at a reference blur of 8 arc min for edges with a band height of 40 arc min are plotted as a function of velocity for observer AP (open circles). For comparison, the prediction by a fully separable blur model with parameters from the fit to the data of the smaller reference blurs is plotted (solid line). The prediction is fairly consistent with the thresholds.

The results of this experiment confirmed that band height affects blur discrimination thresholds. In the study of Watt and Morgan (1983b), the band was 12 arc min in height. If the assumption made above is correct, this means that their thresholds for reference blurs larger than about 3 arc min are degraded by the effect of the small band height. This may well be the reason for the unusual power law exponent of 1.5 which they obtained. Most of the studies for spatial dilation thresholds have, like this study, found a power law with an exponent of 1.0 (i.e. Weber's Law). For example, Campbell, Nachmias and Jukes (1970) found that spatial frequency difference thresholds are a fixed proportion of the criterion frequency. Levi and Klein (1990) reported that spatial interval discrimination thresholds are directly proportional to the separation of the lines.

4.11. Experiment 16: bright end leading

The results of the repeated main experiment suggest that the blur discrimination system can experience low-level learning that can change thresholds differently across reference blurs. The bias experiments, on the other hand, suggested differences in short-time-scale adaptation depending on the contrast polarity of the edge in relation to its direction of motion. Moulden and Begg (1986) demonstrated a simultaneous specificity of adaptation to the direction of motion and the contrast polarity of an edge. This adaptation had a time-scale of minutes. They also discovered a long lasting (many days) direction- and polarity-specific adaptation which they could not explain.

In the main experiments the low-luminance end of the edge had always been the one leading in motion. This experiment was fully identical with the main experiment except that the high-luminance end was leading. The aim was to search for such deviations from the results of the main experiments that could indicate direction- and polarity-specific learning. The observer was the author (AP).

Results and discussion

The results are shown in Fig. 4.25. The gross appearance of blur discrimination thresholds for edges with the high-luminance end leading resembles that for edges with the low-luminance end in front in the main experiment. There is, however, one important difference. For each reference blur in Fig. 4.25, there is a jump in the threshold curve at the same velocity, approximately between 1 and 2 deg/s. For a reference blur of 4 arc min, this jump is towards smaller values; for smaller reference blurs, the jump is upwards. It was noticed that these jumps may be related to changes in threshold between the main and the repeated main experiment. To demonstrate the relation, blur discrimination thresholds for a reference blur of 4 arc min from this and the main experiments were all plotted in the same figure (Fig. 4.26).

Fig. 4.26 shows that the thresholds for stationary and slowly moving edges with the bright end leading are close to those for the same velocities in the repeated main experiment where the dark end was leading. This is an expected finding - a stationary edge has no contrast polarity in relation to the direction of motion. However, at a velocity of 2 deg/s, the threshold drops near the value

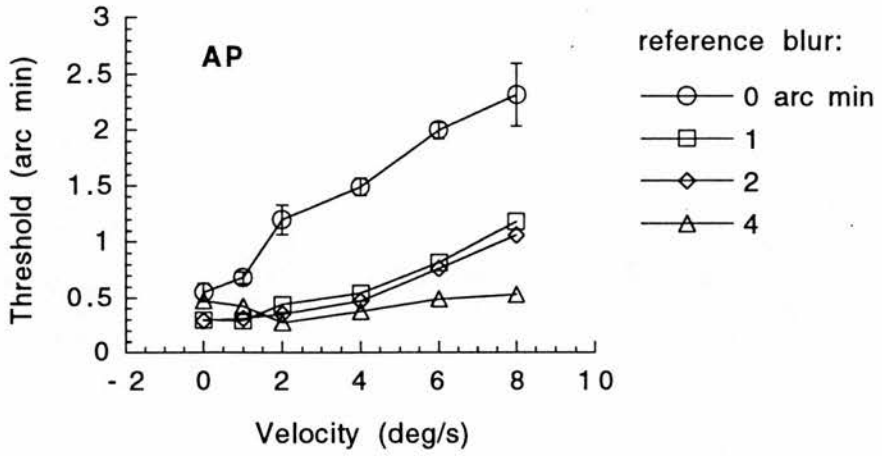


Fig. 4.25. Blur discrimination thresholds for edges having the high-luminance end leading in motion are plotted as a function of velocity at four different reference blurs (space constants 0, 1, 2 and 4 arc min) for observer AP. Error bars for zero arc min reference blur are presented as an example.

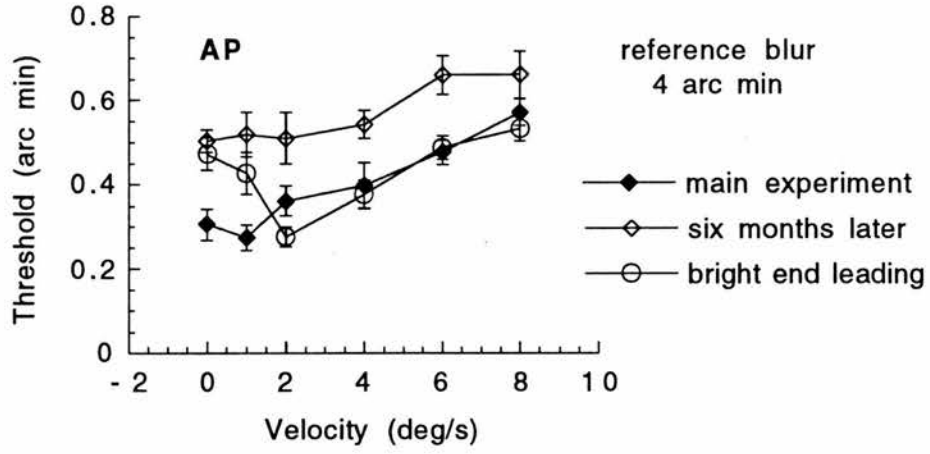


Fig. 4.26. Blur discrimination thresholds for a reference blur of 4 arc min from the main experiment (filled diamonds), from an identical experiment (open diamonds) carried out six months later, and from an otherwise identical experiment except that the high-luminance end of the edge was leading in motion in contrast to the low-luminance end in the other two.

obtained in the original main experiment, and follows closely to the results of that experiment at higher velocities. The changes for other reference blurs were analogue, but less distinct, partly because the differences between the thresholds in the original and the repeated main experiments were smaller. The finding has two important implications. First, it shows that the learning in blur discrimination is specific to the contrast polarity of the direction of motion, or to the direction of change in luminance. The second implication is even more remarkable. Not only does the finding suggest that there are separate systems for spatial analysis of stationary and moving targets, but it also implies that the system for spatial analysis of moving objects can be further divided into two subsystems that are here termed onset and offset systems. However, the results do not rule out the possibility that the stationary and the onset system were the same.

The finding is in accord with the suggestion of Tolhurst (Kulikowski & Tolhurst, 1973; Tolhurst, 1973) that there are two separate systems for motion analysis: movement-sensitive or transient channels and movement-independent or sustained channels. However, it does not support Tolhust's claim that the transient system provides information only about the motion and not about the structure of a moving object, and that the sustained system conveys information only about the spatial structure. Thus, his proposal separated the perception of form and motion of a moving object. The results presented in Fig. 4.26 suggest a different interpretation: there seem to be separate systems for the analysis of the spatial form of stationary and moving objects. This finding agrees well with the results and interpretations of Burr (1981) and Anderson and Burr (1985), though later Burr and his collaborators have argued more for a continuum of spatiotemporal receptive fields or filters tuned to different velocities (Burr & Ross, 1986; Burr, Ross & Morrone, 1986; Burr, 1991).

The fact that the modelling of blur in the main experiments has worked without signs of discontinuity has two possible explanations. First, this result would be obtained if the onset system for moving targets and the stationary system were the same. Second, if the properties of the spatial and temporal filters of the systems were initially the same and their learning would happen at the same rate, no discontinuities would be observed. There are no data to predict which one of these possibilities is more likely.

There are no previous suggestions in the literature of psychophysics that the spatial analysis of moving objects might be carried out by separate onset and offset systems. A distant analogue to this possible separation can be found in the structure of the visual pathway: both the transient and the sustained ganglion cells of primates have their own ON- and OFF-centre systems, and these systems remain distinct through the lateral geniculate nucleus and possibly even further. One important consequence of the finding is that if the interpretations are correct, models for spatial analysis of moving objects need reformulation.

4.12. Summary

In a series of experiments, blur discrimination thresholds were measured as a function of velocity and reference blur. A difference in the perceived blur extent was found between edges having opposite contrast polarities leading in motion. This was taken into account in the design of the main experiment. Thresholds in the main experiment showed that motion produces equivalent spatial blur. The amount of this equivalent blur was estimated by modelling the internal representations of blur using linear concepts and by fitting the models to the data. When the results from the fitting are combined with the corresponding results from the repeated main experiment, they show that the velocity dependence of the equivalent blur produced by motion is linear, and that its amount can be predicted by a temporal impulse response with a standard deviation of about 5 ms in normal room light. This suggests that the blur-producing motion integration time is about 20 to 25 ms - much less than the total temporal integration time that is known to be about 100 ms. The difference cannot be explained without a second integration that does not produce blur. As a function of reference blur, blur discrimination thresholds were found to approach asymptotically to a Weber's Law relation. When the direction of motion was perpendicular to the edge profile, velocity had no discernible effect on the thresholds, suggesting that motion blur results from camera-like summation and not from the use of larger isotropic spatial filters for moving than for stationary objects.

The repeated main experiment revealed changes in thresholds and model parameters, indicating low-level learning in the blur discrimination system. When the leading end of the edge was changed, learning effects were not transferred. This suggests that spatial analysis of moving objects may be served by two separate subsystems that are possibly related to the on and off systems of the visual pathway.

5. Spatial interval discrimination experiments

This chapter presents three experiments under the heading of spatial interval discrimination. A spatial interval discrimination task measures the ability to judge the extent of the separation between a pair of dots or lines for a given base separation. In two of the experiments the base separation is zero, making the tasks essentially two-dot resolution tasks.

The general purpose of the experiments was to confirm and refine some of the findings about spatial filtering and temporal integration of moving targets that were obtained in the blur discrimination experiments.

5.1. General methods

The stimuli were generated by the same apparatus as in the blur discrimination experiments, consisting of a Macintosh IIcx computer, a digital-to-analogue converter card and a Hewlett Packard model 1333a oscilloscope with a brief persistence P15 phosphor. A program written by the author generated the stimuli and carried out the same tasks necessary for the execution of the experiment as described in Section 4.2.1.

Dots and lines were displayed on the oscilloscope screen. Digital x and y position information for each dot were converted to analogue voltage levels by the digital-to-analogue converter card. When a line was needed, it was produced by generating a row of unresolved dots placed 0.3 min of arc apart. The timing of stimulus presentation was controlled by the VIA timers of the computer with an accuracy of about 5 μs. The luminance of each dot or line was controlled by a z modulation voltage provided by a third DAC channel. The luminance values of the dots and lines were estimated in the following manner. A given dot was blurred by adding high-frequency triangle-wave signals to the x and y voltages. This transformed the dot to a rectangular patch. The luminance and the area of the patch were then measured. The area of the dot was defined as the region inside the radius of the spot at the half amplitude level of its luminance. A value of 0.15 mm was used for this radius; for modulation transfer function measurements for HP 1333A, see Morgan and Watt (1982). The luminance of the dot was

then calculated by a simple scaling from the luminance of the patch. One should note that the luminance value obtained this way is about a half of the peak luminance of the dot. For the lines, blurring was needed in only one dimension, but otherwise the principle for estimation of the luminance was the same.

Viewing was binocular with natural pupils from a distance of 100 cm. On each trial, the fixation mark was first displayed for 500 ms. Then there was a blank period before the first pattern appeared. After its brief presentation there was again a blank period before the second pattern was displayed. An adaptive probit estimation (APE) algorithm (Watt & Andrews, 1981) with 64 trials was used to measure the thresholds. The observer's task was to decide whether the separation of the dots or the lines was larger in the first than in the second pattern. In a two-dot resolution task the separation was zero in one of the patterns, and the task was essentially to distinguish a small separation from no separation. Threshold was defined as the standard deviation of the resultant psychometric function (83% correct point). At least three independent thresholds were determined for each condition. Each final value represents the root mean square of these estimates. The standard error is the standard error of the mean of these estimates. The same three observers (MS, RO and AP) as in the blur experiments each participated in one or more of the experiments.

5.2. Experiment 17: two-dot resolution as a function of velocity

The main purpose of this experiment was to confirm the finding that the size of the effective spatial filter does not increase isotropically with velocity. In the blur experiment (Section 4.6) where the motion trajectory was perpendicular to the edge profile, it was found that velocity had no discernible effect on the thresholds over the velocity range of 0 to 8 deg/s. This was regarded as evidence against any isotropic motion filter. However, another explanation related to the finite height of the band was also presented: when a band of 20 arc min in height sweeps vertically over a hypothetical horizontal stripe on the retina, the luminance profile of the stripe stays constant for at least 40 ms for the velocities tested. Thus, filters for stationary objects could be used in spatial analysis. In a two-dot resolution experiment, the target has no extent in the direction of motion when it is perpendicular to the separation. If perpendicular motion did not affect the resolution thresholds, the existence of isotropic motion filters could be ruled out.

Another inspiration to this experiment was a study of Levi and Klein (1990). They found that in the fovea, the two-line resolution threshold and thus the equivalent intrinsic blur (a factor relating these two was found to be close to unity) is about 0.5 arc min. By definition, the equivalent intrinsic blur of Levi and Klein is the same as the effective static spatial filter in the blur discrimination experiments where its best fit estimates for observers with normal vision were found to be consistent with the intrinsic blur values reported by Levi and Klein. This experiment made it possible to study whether a two-dot resolution task gives comparable estimates for the effective static spatial filter to those obtained in the blur discrimination experiment for the same observers. Some interest was also focused on the velocity dependence of two-dot resolution threshold. It was expected that it might be close to that of blur discrimination for a reference blur of zero arc min, which itself is in fact a resolution threshold.

Methods

The stimuli were bright dots on the oscilloscope screen with a background illumination of 25 cd/m^2. This illumination was produced by normal room lighting, ensuring that all observations were carried out at photopic levels. The luminance of each dot was 600 cd/m^2. The separation of the dots was always horizontal. The dot pairs were either stationary or moved horizontally with the same speed randomly either leftward or rightward. In the case where the direction of motion was perpendicular to the separation, the dot pairs moved vertically randomly either leftward or rightward.

Each pair of dots was moved by shifting its position less than 1.44 min of arc in every 3 ms so that the motion appeared to be continuous. The refresh rate was approximately 8000 Hz. To avoid pursuit eye movements, the duration of the presentation of each pair dots was 150 ms, known to be too short for the observer to track the object by smooth eye pursuit. The total number of stations or frames was 50. To minimize the effects of anticipatory eye movements, the durations of the blank periods in each trial were randomized, the one between the fixation mark and the presentation of the first pair of dots in the range of 300 - 500 ms and that between the dot pairs in the range of 450 - 750 ms. The spatial information at the onset and offset of the trajectories was masked by varying randomly (uniform distribution in the range of ± 3 frames) the time that the dots were turned on and off.

At least three thresholds were determined for each condition. Each final value reported represents the root mean square of these estimates. Two observers (MS and AP) measured two-dot resolution thresholds for six different velocities (0, 1, 2, 4, 6, and 8 deg/s) both with the direction of motion parallel and perpendicular to the separation of the dots.

Results and discussion

The results are shown in Fig. 5.1. For both observers, the two-dot resolution thresholds increase approximately linearly with velocity when the separation of dots is along the direction of motion. When the separation and the direction of motion are perpendicular, increasing velocity does not increase thresholds. For both observers, the two-dot resolution threshold for a stationary target is about 0.7 arc min.

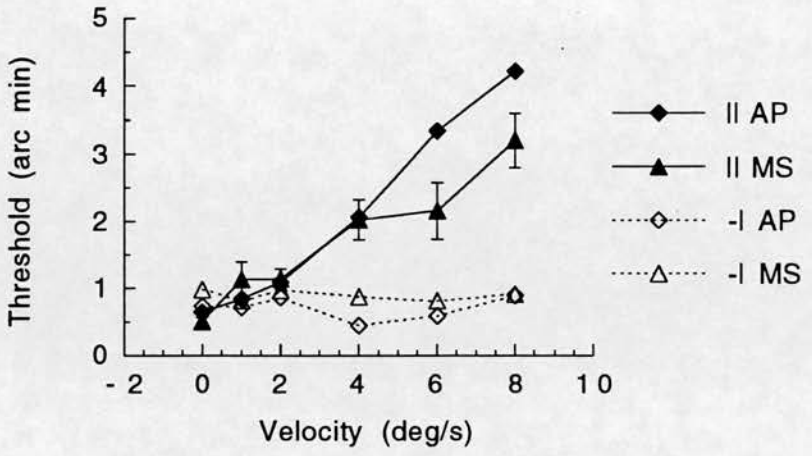


Fig. 5.1. Two-dot separation thresholds for observers AP and MS for dots having their separation along the direction of motion (solid lines, filled symbols) and perpendicular to it (dashed lines, open symbols). Error bars are presented as an example for one observer in one condition - on the average standard errors were less than 10% of the threshold. The most important finding is that velocity does not increase thresholds when motion is perpendicular to separation.

The results confirm that the size of the effective spatial filter does not increase isotropically with velocity. For any isotropic filter, the blurring effects should be similar in all directions of motion. It is possible to explain the results with non-isotropic motion filters with space constants increasing as a function of velocity. However, the space constants should increase only in the direction of motion - the filter profiles perpendicular to the direction of motion should not change with velocity. This seems to be in contrast with the suggestion of Anderson et al. (Anderson & Burr, 1987; Anderson & Burr, 1989; Anderson & Burr, 1991; Anderson, Burr & Morrone, 1991) that motion detectors are as long as they are wide.

The two-line resolution threshold for observer MS was about the same as the estimate for the effective static spatial filter obtained for MS in the blur discrimination experiment. For observer AP, the resolution threshold is about 30% smaller than the filter size estimate. On the average over both observers, the two-line threshold is about 85% of the effective static spatial filter. This is in line with the results of Levi and Klein (1990). They found that in the fovea, the factor relating the two-line resolution

threshold and the equivalent intrinsic blur was close to unity, and in fact it was 0.9. From this it can be concluded that in this experiment, the two-dot resolution task gave estimates for the effective static spatial filter that are close to those obtained in the blur discrimination experiment.

The velocity dependence of two-dot resolution thresholds is much stronger than that of blur discrimination for a reference blur of zero arc min. With linear approximations, the corresponding mean slopes for the two observers are 0.39 and 0.15 arc min/(deg/s). This suggests that these tasks may use different types of mechanisms and cues in discrimination. Conventionally, the resolution task is thought of as discriminating two dots or lines from one. However, Levi and Klein (1990) proposed that as the target in these tasks always appears single at threshold, the most sensitive cue might be the size cue, i.e. the widening of the luminance distribution. This cue is in principle analogous to that of blur discrimination, and it would predict an approximately similar velocity dependence. However, since this dependence is different, the size cue may not be the only cue in two-dot resolution tasks, at least for moving targets.

5.3. Experiment 18: two-dot resolution as a function of exposure duration

One of the most important studies dealing with motion blur is by Burr (1980), which measured the perceived amount of motion smear by matching the length of a short stationary line to the apparent length of dots in motion. The results have been shown earlier in Fig. 2.14 in Section 2.5.6 of this thesis. For exposure durations shorter than about 30 ms, the dots were seen to be smeared approximately to an amount corresponding to the distance travelled. Beyond 30 ms, the perceived smear decreased with increasing durations. Burr, Ross and Morrone (1986) interpreted this to indicate that "the mechanisms responsible for the detection of objects in motion function to remove smear, provided that there is sufficient time for them to operate."

One problem with the experiment of Burr (1980) was the possibility of subjective bias. It is easy to discriminate between a stationary line and a dot sweeping the same length, and thus the observer may for instance continuously underestimate the streak of the moving dot, without realizing it himself. The rather large difference between the results of the two observers in Fig. 2.14 may reflect this bias. To verify the results of Burr using a paradigm without subjective bias, a two-dot resolution task was selected as the method. It was found in the previous experiment, that two-dot resolution thresholds increase approximately linearly with velocity, indicating that it can be used as a measure of intrinsic blur produced by motion.

Methods

The stimuli were bright dots on the oscilloscope screen with a background illumination of 25 cd/m^2 produced by normal room lighting. The luminance of each dot was 1150 cd/m^2. The separation of the two dots was always horizontal. The pairs of dots moved horizontally at a velocity of 4 deg/s randomly either leftward or rightward. Each pair of dots was moved by shifting its position 1.2 min of arc in every 5 ms. The effects of anticipatory eye movements were minimized by the same method of randomizing the durations of the blank periods in each trial as in the previous experiment.

At least four independent thresholds were determined for each condition. Each final value is the root mean square of these

estimates. Two observers (MS and AP) measured two-dot resolution thresholds for seven different exposure durations (10, 20, 30, 40, 50, 75, and 100 ms).

Results and discussion

The results for both observers are shown in Fig. 5.2. For unknown reasons, observer MS's performance in this experiment was surprisingly poor and inconsistent compared to her performance in other experiments. For this reason, the following observations have been made mainly on the basis of the results for observer AP. The two-dot resolution thresholds increase linearly with exposure duration up to 30 ms. For durations longer than 30 ms, the thresholds decrease modestly with increasing duration.

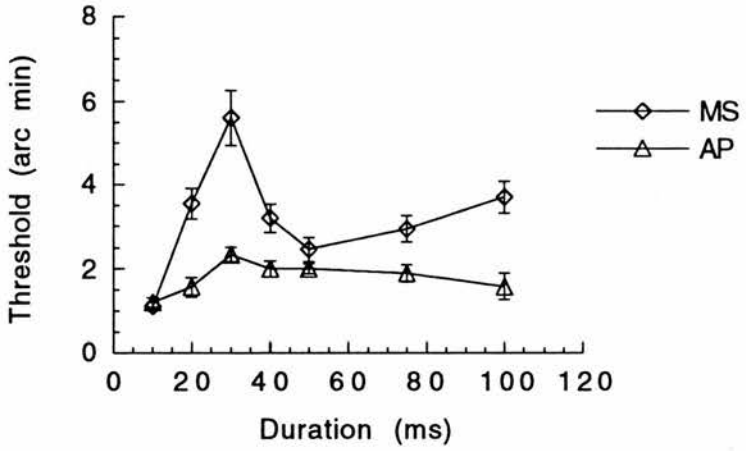


Fig. 5.2. Two-dot separation thresholds as a function of exposure duration for observers AP and MS. The separation of the dots was horizontal, and they moved horizontally at a velocity of 4 deg/s. Error bars represent ± one standard error. The thresholds increase up to a exposure duration of 30 ms, and decrease thereafter.

The dependence of the two-dot resolution thresholds on the exposure duration is similar to that of the perceived streak in the study of Burr (1980). The curve for observer AP in Fig. 5.2 resembles the curves in Fig. 2.14. Thus, the results of this experiment can be seen as a confirmation of the results of Burr. However, differences arise in interpretation. Burr explained the 30 ms turning point by assuming that the mechanisms to remove smear need some time before they start to operate. Here it is proposed that 30 ms is the length of the blur-free integration time. One can

also present other explanations than motion blur removal for the decrease in thresholds and perceived streak lengths after 30 ms.

Watt (1987) demonstrated improvements in sensitivity of several discriminations involving short lines with increased exposure durations, and showed that these improvements can be interpreted in terms of a change of the spatial scale of analysis. The visual system seems to scan from coarse to fine spatial scales over a period of at least one second after the onset of a stimulus. A finer spatial scale would mean less static spatial blur, and thus smaller two-dot resolution thresholds and shorter perceived length of the streak of a moving dot.

Performance can also improve without any changes in static or motion blur. For instance, the data of Watt (1987) for line length discrimination shows that the Weber fraction is the smaller the longer the exposure duration. In two-dot resolution, a decrease in the Weber fraction would mean smaller thresholds. If we assume that the Weber fraction is related to the signal-to-noise ratio, then seeing the streak of a moving dot with greater clarity might also mean as seeing it shorter than an equally blurred streak with a poorer signal-to-noise ratio.

There are even some features in Fig. 2.14 that do not fit to the blur removal explanation. For observer ML, the difference between the curves for velocities of 5 and 15 deg/s is approximately constant for exposure durations from 30 ms upwards. According to the blur removal explanation, the difference at 30 ms corresponds to the difference in motion blur for the two velocities. However, it is against the explanation that none of this difference is removed with increased exposure durations.

5.4. Experiment 19: effects of positional jitter as a function of frame time

The results of the main experiment suggest that the blur-producing motion integration time is about 20 to 25 ms. Since the total temporal integration time is known to be about 100 ms, it reasonable to assume that there exists a second integration stage that does not produce blur. The shifter circuit principle (Anderson & van Essen, 1987) and the spatiotemporal receptive field model (Burr, Ross & Morrone, 1986) propose that motion blur is not produced because the integration follows the path of the moving object. However, some recent studies have suggested that blur-free integration may not be dependent on linear motion.

Badcock and Wong (1990a; 1990b) measured the ability of observers to detect small changes in the separation of two parallel vertical lines. In this spatial interval discrimination experiment, observers were shown a pair of lines for 600 ms, followed by a dark interval of 100 ms and then a second pair for 600 ms. Badcock and Wong found that correlated horizontal jitter of the line pair ranging at least up to 8 arc min had little detrimental effect on performance for two lines with a separation of 6 arc min. The jitter was produced by presenting the line pair in a new randomly chosen location for 3 ms in every 30 ms. Each new location was chosen from a uniform probability distribution extending the horizontal jitter range, i.e. the lines had equal probability of appearing anywhere in the range. In a related experiment, they also found that the performance improves with durations up to at least 300 ms.

The present experiment was designed to find out whether the resistance to positional noise is related to the blur-free integration of moving objects. The results of Badcock and Wong could be explained simply by the notion that since there were no spatial displacements inside the blur-producing integration time of 25 ms, no blur was expected. In this experiment, the effect of positional jitter was studied as a function of time interval between the presentations of the line pair at different locations. The prediction was that if this time interval were made shorter than about 25 ms, blur would start to degrade the performance.

Methods

The experimental setup was not exactly identical to that in the experiment of Badcock and Wong (1990a; 1990b), but there were no details that would be expected to produce differences in the basic results. Each trial consisted of two presentations of targets. First, a pair of parallel vertical lines of 15 arc min in height were presented on the oscilloscope screen with a background illumination of 25 cd/m^2. The base horizontal separation between the lines varied in the range of 6 arc min ± 10% from trial to trial. Each line was produced by generating a row of unresolved dots placed 0.4 arc min apart. A given number of frames were shown for a total time of 600 ms. After a blank interval (with duration random in the range of 450 - 750 ms) another pair of lines was shown with the same number of frames for 600 ms. The observer's task was to report whether the separation of the second pair of lines appeared wider than that of the first.

In each frame, the location of the line pair was updated from a uniform probability distribution extending ± 6 arc min from the mean location, and the line pair was presented for 3 ms. The luminance of each line was 650 cd/m^2.

Two observers (MS and AP) participated in the experiment. Separation discrimination thresholds were measured as a function of the number of frames in the 600 ms presentation. The range was selected so that the interval between the onsets of the 3 ms frames varied from 10 ms to 60 ms. In addition, observer AP measured thresholds for cases where there were only 1, 2 and 4 frames in the 600 ms presentation. Observer MS measured at least three thresholds for each condition, and observer AP at least two. Each final value reported represents the root mean square of these estimates.

Results and discussion

Fig. 5.3 shows separation discrimination thresholds for the two observers against the number of frames. With twenty frames, the thresholds are about 20 arc sec. Even though the jitter in this experiment was higher than in that of Badcock and Wong (1990a; 1990b), the thresholds are very close to the values they obtained, thus confirming their result. However, the thresholds start to rise with increasing number of frames when the number exceeds about 20 to 25 frames. This limit corresponds to an interval of about 25 ms between the onsets of the frames. This result confirms the

prediction that if the time interval between the frames is made shorter than about 25 ms, blur starts to degrade the performance. When the frame number exceeds 40, the thresholds seem to reach an approximately constant level. This is not fully in line with the integration theory - the thresholds should increase further. However, the rate of the increase should decrease with increasing frame number. Thus, the upper flat region in Fig. 5.3 may just be a slight deviation from the true curve produced by the probabilistic nature of the measurements.

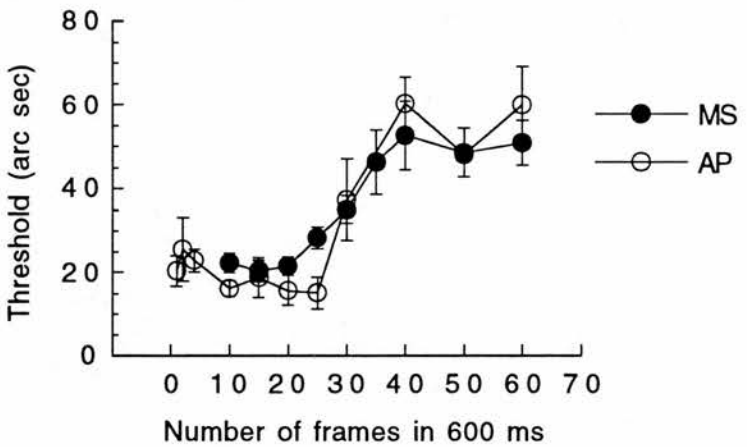


Fig. 5.3. Separation discrimination thresholds for observers AP and MS are plotted as a function of the number of frames in the total 600 ms presentation time of a line pair. The reference separation of the lines was 6 arc min. In each frame, the line pair was shown for 3 ms in a new location chosen randomly within ± 6 arc min of the mean location. Error bars represent ± one standard error. The performance starts to degrade when the number of frames exceeds about 20 to 25, i.e. when the interval between the frames falls below about 30 to 24 ms, respectively.

Badcock and Wong (1990a; 1990b) argued that the importance of the resistance to positional noise in human vision is that it eliminates the effects of eye movements produced by oculomotor jitter. The present results confirm that there really exists resistance to positional noise, but they also show that this resistance has a limitation: jitter that occurs inside a period of about 25 ms degrades performance. The fact that the same 25 ms limit seems to govern visual blur formation both for linear and random motion suggests that resistance to positional noise may be more important than Badcock and Wong suggested. It may reflect the existence of a mechanism responsible for the analysis of spatial form of objects,

no matter what kind of translational visual motion the objects are experiencing. This idea is not consistent with the models of motion-deblurring based on motion-tuned mechanisms (Burr, Ross & Morrone, 1986) or linear shifting (Anderson & van Essen, 1987).

5.5. Summary

Three experiments from the general class of spatial interval discriminations were carried out in order to extend and verify some of findings of the blur discrimination experiments. In the first experiment, two-dot resolution thresholds were found to be strongly dependent on velocity when the separation of the dots was in the same direction as their motion. With perpendicular motion, velocity had no effect on the thresholds, providing evidence for the suggestion that motion blur is caused by camera-like summation and not by a size increase of isotropic spatial filters with velocity.

In the second experiment, the dependence of two-dot resolution thresholds on exposure duration was found to be similar to that of the perceived streak of a moving dot in the study of Burr (1980). The thresholds increased up to a exposure duration of 30 ms, and decreased thereafter. However, it was shown that there exist plausible explanations for this decrease other than motion blur removal.

In the third experiment, the effect of positional jitter was studied as a function of time interval between the stationary frames of a jittering target. The results confirmed the finding of Badcock and Wong (1990a; 1990b) that the visual system has resistance to positional noise. However, an important limitation was also found: jitter that occurs inside a period of about 25 ms degrades performance.

Evidence that the blur-producing integration time is about 25 ms has accumulated throughout these experiments. For consistency with the total temporal integration time known to be about 100 ms, a second integration stage that does not produce blur must be assumed. The next chapter introduces a theoretical alternative to this integration - a solution that would form a simultaneous representation of motion, form and location of an object.

6. Displacement mapping - a global model for visual motion processing

6.1. Introduction

Motion-deblurring cannot be regarded as a separate entity in motion processing. It is easy to agree with Burr, Ross and Morrone (1986) that deblurring is an integral part of the general mechanism responsible for the analysis of the spatial form of moving objects. It is not as self-evident that the same mechanism would also be responsible for detection of velocity as the spatiotemporal receptive field model of Burr et al. proposes. That model is motion-tuned, but the results of Badcock and Wong (1990a; 1990b) and those of the positional jitter experiment of this thesis suggest that motion-tuning might be an unnecessary restriction for the mechanism.

This chapter proposes a model that can simultaneously derive information about shape, motion and location of moving objects without motion-tuning. The basic principle is presented together with an algorithm and some selected psychophysical and neurophysiological implications of the model. The aim of this proposal is not to provide a general model to solve all the problems of the theories of visual motion processing, but to inspire discussion and to point at some connections which may have not been noticed before.

6.2. Basic principle

The overall analysis of motion can be divided into two levels: first, the stages responsible for the measurement of motion in a changing two-dimensional image, and second, the stages using the motion information. How this information can be used depends on how the measurement stages present it to the later stages. Much of the research and theories so far have concentrated on the measurement and computation of the two-dimensional velocity field of the image, i.e. on the computation of optic flow. In these theories a velocity value is calculated for each point in the image, and only local motion measurement values are used in this calculation. It is

generally assumed that this kind of topographical velocity representation is essential in recovering the three-dimensional structure of objects from motion. However, there may be other visual tasks that would benefit more from a different type of velocity representation.

This chapter presents an alternative approach to the presentation of motion information. The basic principle is not entirely new, but the algorithm has not been presented before. The approach is called here a displacement mapping concept. A point $(\Delta x, \Delta y)$ in a displacement map corresponds to orthogonal displacements Δx and Δy in one frame relative to the next. The value of a point $(\Delta x, \Delta y)$ shows the total amount of displacements with a Δx and Δy shift between the two consecutive images. The concept is easier to understand if one regards the displacement map as a two-dimensional histogram of displacements. If the time interval between the two images is the same for each point, the displacement map is also a velocity map. Frames are used here to describe the principle in theoretical and algorithmic levels, but this does not mean that they would be real entities of the implementation.

The displacement or velocity map is in principle identical to the motion map by Barlow (1981) who proposed it as one example of a more general concept of non-topographical mapping. Being non-topographical is one of the most important properties of a displacement map. Barlow also gave a good description of the difference between a topographical and a non-topographical map:

"In a topographical map neighbourhood relations are preserved; if two points lie near each other in the original, they will lie near each other in the map. In a non-topographical map it is not topographical propinquity that is preserved, but propinquity along some other dimension such as movement or colour, and in such a map points moving similarly or coloured similarly will lie close to each other, no matter how far apart they were topographically in the original."

6.3. An algorithm to calculate a displacement map

In order to compare the spatial relations between two images one has first to decide the type of input values to the comparison operator. To simplify the discussion, the initial light intensity measurements of the photoreceptors have been chosen here as motion primitives, but some more symbolic primitives, such as zero-crossings or centroids, might be equally adequate. The most natural

choice for the comparison operation is the two-dimensional cross-correlation of the images. The maximum in the cross-correlation function of two images or frames shows the optimum match or the optimum displacement between the frames. This correlation function, however, is not just a description of the displacements between frames, but it also contains information on the positional relations of those points which have not changed their position between the frames. A simple way to remove most of this part is to subtract the autocorrelation function of the second image from the cross-correlation function. If we use $f_1(x,y)$ to denote the first frame and $f_2(x,y)$ the second frame, the displacement map can be defined as

$$D(\Delta x,\Delta y) = f_1(x,y) \otimes f_2(x,y) - f_2(x,y) \otimes f_2(x,y) \qquad (6.1)$$

where $\otimes$ denotes correlation operator. This equation can be reduced to a form

$$D(\Delta x,\Delta y) = [f_1(x,y) - f_2(x,y)] \otimes f_2(x,y) \qquad (6.2)$$

This algorithm is one possible mathematical solution to the calculation of a displacement map. There is no evidence that real neural networks would do it like this if they were calculating displacement maps. However, this algorithm contains components that are possible to implement using neural hardware. As such the algorithm suffers from the same weakness as for instance the minimal mapping theory proposed by Ullman (1979) for the computational solution of motion correspondence problem. This weakness is the combinatorial explosion caused by the need that for the correlation process, each point in one image should be connected with all points in the other image. A solution to this problem is to cut the image into combinatorially smaller pieces. This could probably be done without losing the advantages of the concept by using different spatial scales: a large patch of the visual field could be scanned with a coarse spatial scale, then for instance each quarter of the patch with a scale of half of that of the patch, and so on. In each part the number of elements would have to be sufficiently low to make the neural wiring possible. This solution would lead to a number of displacement maps with different sizes and different spatial scales. Since a displacement map is translation-invariant, the final map could be formed simply by superimposing the submaps

with a proper intensity weighting. This way, the final map would be inhomogeneous in spatial scale: the closer the origin, the finer the scale. Another way of thinking to produce the same map would be to assume that the correlation process itself is performed inhomogeneously, i.e. the further two points or neurones lie from each other, the lower the probability that they are connected. This kind of connectivity, on the other hand, is a property of cortical neurones (Douglas & Martin, 1991).

One-dimensional elaboration of the algorithm

Let us elaborate Eq. (6.2) further in the case of a single moving object against a stationary background. To simplify the treatment, consider a one-dimensional image formed by discrete points representing, for instance, activities of photoreceptors or ganglion cells. Let the first frame consist of a background $B(i)$ and an object $O(i)$. If we ignore the occlusion of the background by the object, we get

$$f_1(i) = B(i) + O(i) \tag{6.3}$$

If the object moves a distance v in the direction of the positive i-axis between the frames, we get for the second frame:

$$f_2(i) = B(i) + O(i - v) \tag{6.4}$$

Substituting Eqs. (6.3) and (6.4) into Eq. (6.2) yields

$$D(\Delta i) = [O(i) - O(i - v)] \otimes [B(i) + O(i - v)] \tag{6.5}$$

If the borders of the receptive field of the displacement mapping mechanism are $-R$ and R, then

$$D(\Delta i) = \sum_{i=-R}^{R} [O(i) - O(i - v)][B(i + \Delta i) + O(i - v + \Delta i)] \tag{6.6}$$

This can further be rewritten as

$$\begin{aligned} D(\Delta i) = {} & \sum_{i=-R}^{R} O(i)\, B(i + \Delta i) + \sum_{i=-R}^{R} O(i)\, O(i - v + \Delta i) \\ & - \sum_{i=-R}^{R} O(i - v)\, B(i + \Delta i) - \sum_{i=-R}^{R} O(i - v)\, O(i - v + \Delta i) \end{aligned} \tag{6.7}$$

This is the equation for a displacement map in a one-dimensional image where the background does not move between the frames. To clarify the properties of this equation let us consider two special cases.

Point object

It is very common in natural scenes that the image of the moving object is only a small part of the whole scene. For instance, when we see a bird flying or the tip of a pen sliding on the paper when we are writing, the object is, in practice, a point object. Let us define a point object which has a unit luminance and which is placed in position s in the first frame:

$$O(i) = \begin{cases} 1, & when\ i = s \\ 0, & otherwise \end{cases} \qquad (6.8)$$

When we substitute this in (6.7), only that term in each of the four summations is preserved where the index of the first O term is s. When we solve for i in each case of the first O term and substitute the result in the second term, we obtain

$$D(\Delta i) = B(\Delta i + s) + O(v) - B(\Delta i + v + s) - O(0) \qquad (6.9)$$

This equation shows that the point object maps to location v on the positive side of the displacement map, and at zero on the negative side. The sides are assumed to be separate on the basis of the possible implementation: neurones cannot represent negative activities. On both sides, the background is shifted so that the spatial relations of the object and the background are the same as in the second frame. Motion can be detected from the map simply by checking if there is any activity at location zero on the negative side of the map. The moving point can be identified by looking at the same location. The co-ordinate of the point on the positive side gives its velocity, and its location in the background is correctly represented on both sides of the map. Fig (6.1) illustrates these principles in a case where a point object moves amongst three stationary point objects.

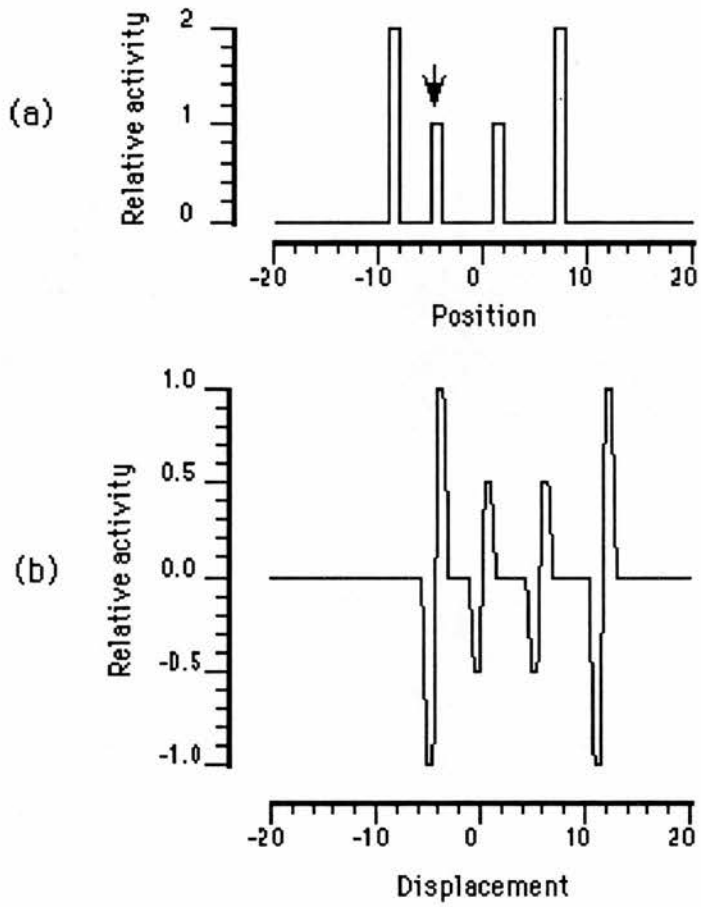


Figure 6.1. An example of displacement mapping in the case of one-dimensional point object. (a) The first frame. The moving object is marked with an arrow. The second frame which is not plotted is similar to (a) except that the object has moved 1 unit to the right. (b) The displacement map. The map simultaneously identifies the moving object, displays its position in the background, and shows its displacement (or velocity) between the frames.

Object and background uncorrelated

If the moving object is more than just a point, the displacement map is not as simple as Fig. (6.1). If we assume that the spatial information of a larger object and that of the background are uncorrelated, the first and the third term of Eq. (6.7) approach zero and the equation reduces to

$$D(\Delta i) = \sum_{i=-R}^{R} O(i)O(i-v+\Delta i) - \sum_{i=-R}^{R} O(i-v)O(i-v+\Delta i) \qquad (6.10)$$

The positive term of this equation is equivalent to the autocorrelation of $O(i)$ with its origin shifted to location v. The negative term is equivalent to the autocorrelation without any shift. Thus, on both sides of the map the presentations of the object are located identically to the case of the point object. Also in this case, the map can be used to detect the motion, to identify the moving object and to define its velocity. The spatial information about the shape of the object is presented in the form of its autocorrelation. The map does not contain information about the location of the moving object in the background.

Two-dimensional illustration of a displacement map

One-dimensional examples serve to clarify some of the properties of displacement mapping, but a two-dimensional example will illustrate better what they look like. One of the most frequently used test patterns for models of motion processing has been the sine-wave grating. In this example, a vertical sine-wave grating moves horizontally to the right about a third of a cycle between two consecutive frames. The example is illustrated in Fig. 6.2; the first frame in (a), the second frame in (b), and the resulting displacement map in (c). The grey levels in (c) have been adjusted for presentation purposes; that of the surrounding area corresponds to an activity level of zero, areas darker than that correspond to the positive side of the map, and lighter to the negative side. There are several local maxima in the displacement map, each corresponding to a physically plausible displacement of the sine wave between the frames. However, there is only one global maximum, and its position corresponds to the minimum displacement that can explain the change from frame one to frame two. This is consistent with the human percept in a case like this. The existence of the global maximum is further illustrated in (d) where the displacement map of (c) has been thresholded to leave only the maximum visible. The mathematical reason that the correlation of sine waves produces a global maximum is the restricted field of view or the aperture - the result reflects the fact that when the data window is rectangular, the correlation window has a form of a triangle.

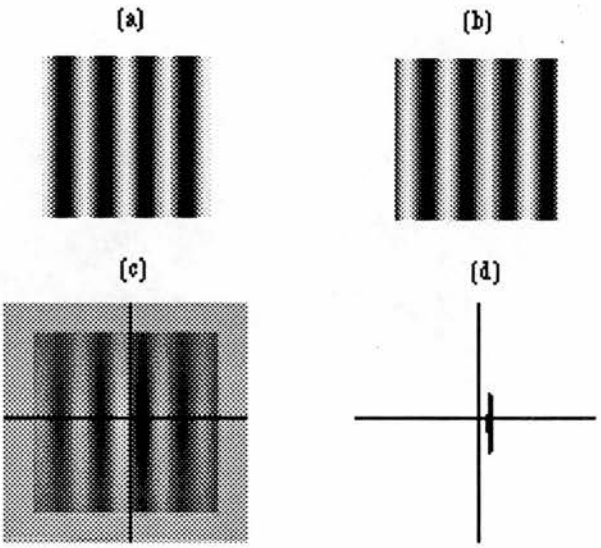


Fig. 6.2. Displacement mapping for a sine wave grating. (a) The first frame. (b) The second frame where the grating is displaced about a third of a cycle to the right relative to the first frame. (c) The resulting displacement map. The grey level of the surrounding indicates zero activity. The displacement map is periodic, but there is a global maximum. Its position corresponds to the displacement between the frames. In (d) the map has been thresholded to emphasize the global maximum.

6.4. Comparison with other models of motion detection

Two approaches to detecting motion can be distinguished in the literature: correlation and gradient models. The basic mechanism of the correlation models (Reichardt, 1957; van Santen & Sperling, 1984; Adelson & Bergen, 1985; Watson & Ahumada, 1985) consists of two spatially separated input units connected to a multiplier unit, one directly and the other through a delay. The detector responds when a stimulus moves from a point seen by the delayed unit to a point seen by the direct unit and the time taken to the move equals the delay. The general principle of the gradient scheme is that the velocity of a given point in an image can be obtained by dividing the temporal derivative of the light intensity at that point by the spatial derivative at the same point.

Both the correlation and the gradient models have their velocity estimation based on local measurements by elementary detectors. The outputs of the elementary detectors are assumed to be explicit, and requirements have to be set for the properties of the detectors to make each of them able to derive the correct velocity. For an example of this, see the elaborations (van Santen & Sperling, 1984) made to the original Reichardt model to prevent spatial and temporal aliasing. In this respect, the displacement mapping principle is fundamentally different from the conventional correlation and gradient models. It is a global model with no elementary detectors having explicit outputs. The correct velocity is defined by the global maximum in the displacement map, as shown in Fig. 6.2. Aliasing effects can be present in the map but no special requirements are needed to exclude them from the motion percept.

Being global may be a more important property of motion processing than what has been previously thought. The local models cannot account for many fundamental psychophysical findings, e.g. for motion capture. This has recently led to theoretical elaborations of the local models which achieve better results by integrating information about motion over local spatial neighbourhoods (Horn & Schunk, 1981; Yuille & Grzywacz, 1988; Bülthoff, Little & Poggio, 1989; Bischof & Di Lollo, 1990). The central idea of these models is to introduce an additional smoothness constraint in the calculation of flow field, and this inevitably requires integration over area. Compared to these integration schemes, the displacement map calculation differs in that at least in principle, each point is influenced by the whole image, and not just by neighbouring points.

The displacement map calculation has both correlation and gradient-scheme properties embedded in it. The result of the correlation of the first frame with the second can be seen as an output field of a large number of simple correlation detectors where the outputs of detectors having the same sampling base and orientation are added together. The time difference between the frames corresponds to the delay unit of a correlation detector. On the other hand, the subtraction of the frames in Eq. (6.2) results in a sort of temporal derivative, at least if we assume that the delay is small.

6.5. Motion correspondence

The perception of motion by the human visual system does not require that objects move continuously across the visual field. Motion can be perceived even when the objects are presented discretely at positions separated by up to several degrees of visual angle. It is a general assumption that this long-range motion phenomenon illustrates the ability of the human visual system to derive a correspondence between the elements in the changing image. From this point of view one can also regard the displacement mapping as a method of solving the correspondence problem. A maximum in the map corresponds to a moving object in the scene. However, strictly speaking displacement mapping does not solve the correspondence problem since it does not try to find the best possible match between the moving points in two images. It is questionable whether the solution of the motion correspondence problem is at all necessary if one reliably measures the velocities of moving objects without it.

6.6. Motion psychophysics and displacement mapping

Short- and long-range processes

Many studies suggest that there are two processes mediating motion perception in the human visual system. This distinction was first proposed by Braddick (1974) who named them the short-range and long-range processes. The short-range process may analyze continuous motion, or motion presented discretely but with only small displacements in space and time. The long-range motion is thought to analyze larger spatial and temporal displacements, as in phi motion. The proposed use of a number of displacement maps with different sizes and different spatial scales in the calculation of the final map may at first seem to be consistent with spatially-limited processes. However, to produce the distinction, the motion system would have to use the intermediate maps separately and if it did, one would certainly expect more than two different processes. In fact, Baker and Braddick (1985) have presented a redefined version of the two-range explanation which has different short-range sub-systems with different retinal eccentricities.

It can be inferred from the results of Nakayama and Silverman (1985) that the upper limit of the short-range motion system's ability to detect discrete displacements (D_{max}) has an inverse linear correlation with the spatial frequency in the random dot field. This is in line with other findings (Chang & Julesz, 1983; Cavanagh, Boeglin & Favreau, 1985; Bischof & Di Lollo, 1990; Cleary, 1990). Phi motion does not have the same kind of restriction: phi movement can be seen over much greater distances. This phenomenon has been thought to be mediated by the long-range motion process.

There is, however, a marked difference in the stimulus between phi motion and random dot displays. A phi-motion stimulus has two frames where a limited number of objects change places, but no new object appear or any old disappear. In a random dot display, the shift makes some of the dots disappear and wrap around so that they reappear at the opposite end. Globally these dots experience a large displacement in the direction opposite to that of the other points. Locally some dots on the leading edge disappear, and some appear on the trailing edge. In the calculation of a displacement map both the local and global effects introduce noise in the form of irrelevant displacements. The amount of this noise depends on the number of dots wrapped around in the display. In a constant shift this number further depends on the spatial frequencies in the display. This shows that it might be possible to explain the D_{max} phenomenon in terms of noise behaviour in displacement mapping.

Motion capture

Unambiguous motion signals from certain coarse image features may seem to inhibit signals from finer details. Ramachandran and Inada (1985) alternated two uncorrelated random-dot frames to generate incoherent noise. When they superimposed a sine-wave grating of low spatial frequency compared to the dots on the display and displaced the grating horizontally between the frames, they found that all the dots in the display appeared to jump along the grating. Ramachandran`s explanation for this motion phenomenon is that the spurious motion signals from the dots are inhibited by the unambiguous motion signal from the grating, and since the dots do not appear stationary, the grating motion signal is spontaneously attributed to the dots.

According to Ramachandran (1990), motion capture is a trick to solve the motion correspondence problem. In his explanation all the

motion information of the dots is discarded, and in the case of random dots there is no real information to lose. Our visual system, however, does not have a priori knowledge of that. Even though motion capture may seem perceptually strange, it is not a great problem for computational algorithms. In fact, Bülthoff et al. (1989) argue that any algorithm that integrates information about motion over local spatial neighbourhoods predicts phenomena such as motion capture. In the displacement map the motion signal of the grating produces the global maximum. In the motion of the random dots, all possible velocities are represented, also that of the grating. Thus, the motion signals from the dots just form a rather flat background all over the map, and also at the position of the global maximum. This maximum defines the velocity of the grating, but if there is no restrictions for the motion state of the dots, the maximum also gives the most probable velocity for the dots. Thus, the velocity of the grating is attributed to the dots, and no inhibition is necessarily needed.

Phi motion

If two images presenting objects at different spatial locations are rapidly alternated, one has the impression that objects move smoothly from one location to the other. In some cases the perception of this illusory impression called phi motion is ambiguous. For instance, if the images contain two similar objects both changing places like in the frames in Fig. 6.3, the motion system has to make a decision between possible percepts. In this case the display is bistable in the sense that one can see either vertical or horizontal phi motion. In addition, these two percepts are mutually exclusive, i.e. only one percept can be seen at a time. There is no preference between the percepts, and so the probability of seeing either of them is 50%. Also the displacement map presentation would be totally symmetric.

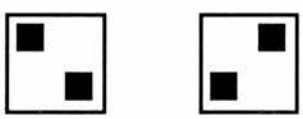


Fig. 6.3. Two frames of a bistable phi motion display.

If a vertical line is placed in the middle of the display as in Fig. 6.4, the percept changes. One sees vertical motion with a much higher probability than horizontal motion. At the start of presentation only vertical motion can be seen. With a prolonged presentation the probability of seeing horizontal motion increases. One can also change the percept by 'deciding' which motion to see.

Fig. 6.4. Two frames of a bistable motion display with preferred vertical motion.

The displacement map corresponding to slightly blurred images in Fig. 6.4 is presented in Fig. 6.5 (a). One can see four maxima (dark patches) in the map. The white blob at the origin of the map shows activity on the negative side, reflecting the existence of motion between the frames. The other two white patches are a consequence of the subtraction of the autocorrelation of the second frame, giving information about the spatial relations of the squares in the second frame. The four maxima correspond to the possible displacements of the squares between the frames: up, down, left, and right. The heights of the maxima are equal, and so that property does not give preference to either vertical or horizontal motion. However, there is something else in the geography of the map that shows a difference between vertical and horizontal motions. The presentation in Fig. 6.5 (b) is a thresholded image of the displacement map. All the points having a value above the threshold are presented as black and all the others as white. There are ridges attached to the vertical motion peaks at a higher level than to the horizontal motion peaks. If the decision about motion is made using, for instance, an integration over area above a certain level and giving preference to the higher value, then vertical motion would be selected.

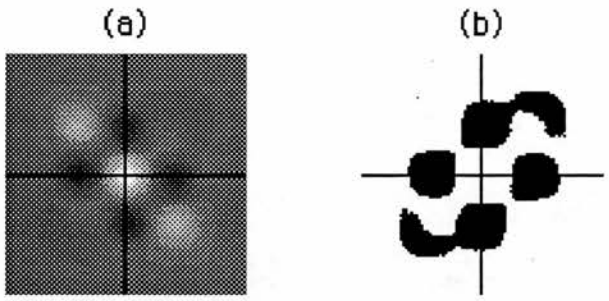


Fig. 6.5. (a) Displacement map corresponding to frames in Fig. 6.4 (with slight blurring). (b) The thresholded image of (a) emphasizes the ridges that attach to the blobs corresponding to vertical motion at a higher level than to those corresponding to horizontal motion.

The increasing probability of seeing horizontal motion when one continues to fixate to the display may be an adaptation effect. The displacement mapping explanation for the vertical motion preference, i.e. the asymmetry of the ridges in the map, is based on the effect of the stationary vertical line in the middle of the display. It is known that the percept of a stationary object decays with time. In this case, this may lead to more equal probabilities between the percepts of vertical and horizontal motion. Another possible stage for adaptation could be the decision process. It could first select the most probable motion, but if the display continues, also bring the other alternative motion of the display into one's consciousness. The displacement mapping concept also offers an interesting explanation for the decision-guided phi motion. One could just focus attention on another maximum in the map, and this would bring out the alternative percept.

Plaid experiments

The visual system must often decide whether the motion signals arise from a single object or from multiple objects. A special case of this problem arises when two drifting gratings with different orientations are superimposed to form a plaid pattern. The observer may see a rigid coherent object (pattern motion) or two gratings moving independently (component motion). Which one is perceived depends on the relative contrasts, spatial frequencies and directions of motion of the gratings. Psychophysical results from plaid experiments have been the inspiration of the two-stage motion

analysis model (Adelson & Movshon, 1982), which is at the moment widely accepted. According to this model, the plaids are first decomposed into oriented, moving gratings by the first stage and the grating speeds are used in the second stage to calculate the plaid speed.

A very interesting plaid experiment is presented by Stoner, Albright and Ramachandran (1990). They found that there is a range of intersection luminances with a high probability of seeing component motion. The midpoint of this range closely coincides the centre of the transparency zone compatible with physics. Thus, the tendency to see component motion depends strongly on whether or not the gratings look transparent. Stoner et al. concluded that motion detecting mechanisms of the visual system must have access to tacit knowledge of the physics of transparency.

The displacement map of a plaid pattern contains three major maxima, one for each of the components and one for the pattern motion. The relative heights and extents of the maxima depend on the relative contrasts, spatial frequencies and directions of motion of the gratings, and also on the luminances of the intersection points. The simplest explanation for different percepts is that the most probable percept is automatically selected by general thresholding and integration rules. In this explanation the tacit knowledge of physics is stored in the connections of the neural network responsible for the calculation of the displacement map. Another explanation could be that the displacement map pattern of the grating is compared to motion patterns in the memory, and the known properties of the best fitting pattern are then attributed to the grating motion.

6.7. Physiological plausibility

The algorithm presented in Eq. (6.2) contains components that may find counterparts in biological hardware. The subtraction of the frames is a practical method of obtaining a time derivative, and based on the experimental findings of Rodieck and Stone (1965) and Dreher and Anderson (1973), Marr and Ullman (1981) showed that the responses of retinal and geniculate transient cells are in close agreement with the time derivative of the Laplacian of a Gaussian filtered intensity distribution on the retina. The second term of the cross-correlation in Eq. (6.2) could be represented by sustained

cells. The separate ON- and OFF-pathways of the visual system could be used in the calculation and representation of the negative and positive sides of a displacement map.

The neurones that could represent the final map are required to be velocity-tuned. Such neurones has been found in the magnocellular stream of the primate visual system: the midtemporal cortical area MT of the macaque monkey has a large proportion of cells whose spatial and temporal tuning covary so as to maintain a constant preferred speed over a broad range of spatial and temporal frequencies (Maunsell & van Essen, 1983). The magno system seems also to be wired up in such a way that it tends to integrate information from different parts of visual field; it seems to able to correlate and link spatially separate stimuli (Livingstone & Hubel, 1987). This is consistent with the global calculation principle of displacement mapping.

A very interesting point is that at the cortical stages, nearly all magno system cells are also orientation-selective. Some researchers have found this a problem in explaining how these neurones can signal visual motion (Gizzi, Katz, Schumer & Movshon, 1990, for example). With the displacement mapping, however, it is not a problem. In fact, orientation-selective behaviour is a property of neurones in a displacement map. Fig. 6.6 presents the displacement map for an oblique bar moving horizontally to the right. The true physical displacement corresponds to the maximum, which also coincides with the centroid of the dark blob. The dark blob shows all the possible displacements in the move. The white blob at the origin corresponds to the subtracted autocorrelation of the bar. Suppose now that a neural population, for instance in the midtemporal cortical area MT, is presenting this map. The tip of the recording electrode is placed in a cell, which is marked in the map with a white arrow. Since the cell belongs to the dark blob, it will respond to an oblique bar moving horizontally to the right. However, when the bar is turned to the vertical position and the test is repeated with the same direction and velocity, the cell does not respond. This is because the dark blob in the displacement map will also turn to vertical position, and thus away from the electrode. Taken together, a neurone representing a point in the displacement map will show both orientation-selective, direction-selective and spatiotemporal tuning behaviour at the same time - like real neurones in the magnocellular system do.

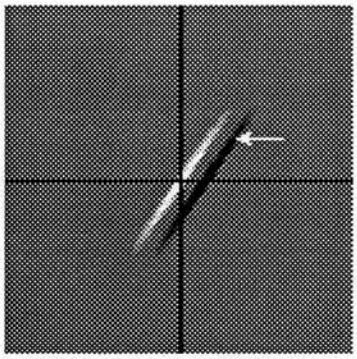


Fig. 4. Displacement map for an oblique bar moving horizontally to the right.

Some indirect evidence for the existence of displacement mapping principle has come from studies of stereopsis. Poggio (1984) found that some cells in the cortical areas of V1 and V2 of macaque monkeys respond to disparity in complex random-dot stereograms. The output of these cells represents the displacements between two images from different eyes and is evidence for the existence of a network capable of making this kind of calculations.

6.8. Discussions on some patterns of motion

It is important for us to know whether an approaching object is going to hit us or pass by. Both the positive and the negative sides of the displacement map pattern of an approaching target have the form of a closed contour, similar in shape to the two-dimensional retinal image of the target itself. If the target is going to hit, the origin of the map is inside the contour. If a large target is approaching us at a high speed, a large area in the map has high intensity values. These properties might be useful, for instance in initiating reflexes to evade danger.

It is generally assumed that the human visual system is able to exploit the intrinsic constraints of rigid motion. This would not be possible, if there were not any well-defined property of motion pattern uniquely associated with this particular class of motion. It is difficult to see whether the displacement map presentation would bring any improvements in the explanations of the detection of

rigid motion. This task may be one of those that needs some other kind of presentation of motion information.

Johansson (1975) showed that a brief observation of patterns of moving lights generated by human figures moving in the dark can lead to a perception of three-dimensional motion and structure of the figures. Displays presenting this kind of motion are called biological motion displays. In some cases it possible to identify the motion just from a few frames, the minimum of course being two frames. Even when the observer cannot immediately identify the motion, he can very quickly say, whether the motion seems natural or not. Our ability to perfoms fast identification of motion patterns may be a consequence of the fact that the displacement maps are the representations of motion patterns that we use to store motion information in our memory.

6.9. Summary

A displacement mapping principle was presented as a way representing of motion information. This method can, without being motion-tuned, simultaneously derive information about shape and motion, and in some cases about location of moving objects. The concept can be used to explain several phenomena of motion psychophysics, for instance motion capture and phi motion. The model is a one-stage model in the sense that all the velocities are calculated in the same stage. One important finding is that the displacement map elements are consistent with the orientation-selective, direction-selective and spatiotemporally tuned neurones in the primate visual cortex.

7. General discussion

The experiments in this thesis have shown that moving objects are blurred to a degree that corresponds to an integration time of about 25 ms. On the other hand, the detectability of moving objects as a function of exposure time suggests that the total temporal integration time is about 100 ms (Burr, 1981). If we assume that the detection of a moving object is made by the same mechanism that is responsible for its spatial analysis, then an integration stage that does not produce blur should also be assumed. This chapter discusses the motion-deblurring models presented in the literature, i.e. the time-space or spatiotemporal receptive field model of Burr, Ross and Morrone (1986) and the linear shifter circuit concept of Anderson and van Essen (1987), in the light of the findings of this study. In addition, an alternative model for temporal integration that is based on similar mathematical operations to those of displacement mapping is proposed.

7.1. Spatiotemporal receptive fields

Burr et al. (1986) derived the spatiotemporal receptive fields by inverse Fourier transform of the spatiotemporal tuning functions constructed using masking results. According to Burr et al., motion smear is determined by the width of the central region of the spatiotemporal receptive field, and by the interaction of several motion detectors. They also stated that in principle, definition as precise as desired may be obtained by the cooperative action of many fields of different profile, but did not propose any mechanism for the cooperation.

The results of the blur discrimination experiments of this thesis are consistent with the view that the apparent spatial filter at each velocity is the result of the convolution between the static spatial filter and the spatial filter determined by the temporal impulse response and the velocity. In the case where the static and the velocity dependent components are fully separable, the space constant of the apparent spatial filter s_a follows the equation:

$$s_a = \sqrt{s^2 + (fv)^2} \qquad (7.1)$$

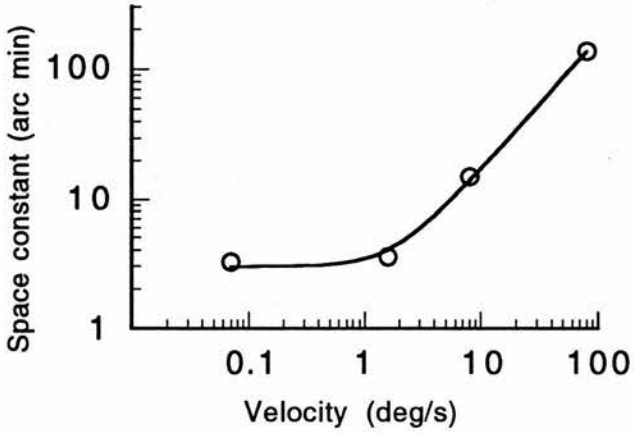


Fig. 7.1. The estimated space constants extracted from the graphic plots of the spatiotemporal receptive fields by Burr et al.(1986) (open circles) are fitted by a model that assumes that the apparent spatial filter at each velocity is the result of the convolution between the effective static spatial filter and the spatial filter determined by the temporal impulse response and the velocity. The good fit indicates that the broadening of the receptive fields with velocity can be explained by motion blur produced by a constant integration time.

where s is the space constant of the static spatial filter, f is the standard deviation of the temporal impulse response, and v is the velocity. To see whether this equation applies also to the spatiotemporal receptive fields of Burr et al. (1986), the apparent space constants of the fields were estimated from the graphic plots they presented for spatiotemporal receptive fields tuned to velocities of 0, 1.6 and 8 and 80 deg/s (redrawn in this thesis as Fig. 2.16 in Section 2.6.1). This estimation was done by measuring the width of the positive central region at t = 0, and by assuming that this one-dimensional profile has approximately the form of the second derivative of a Gaussian. Fig. 7.1 shows the fit of Eq. (7.1) to the extracted values. The fit is good, indicating that the broadening of the receptive fields with velocity may result from motion blur produced by a constant integration time. The fitted parameters, however, are different from the results in the blur discrimination experiment. The space constant of the assumed static spatial filter is about 3 arc min, over four times the mean value

obtained for the observers with normal vision in the blur experiment. One way to explain this difference is to assume that the blur discrimination process uses the smallest spatial filter available while detecting the direction of drift of a sinusoidal grating at threshold uses a much larger effective spatial filter. The difference in the temporal parameter is more important. The estimate for the standard deviation of the temporal weighting function is 28 ms, making the temporal integration time of about 100 ms. This implies that in the psychophysical task of Burr et al. (1986), the whole summation time produces blur, in contrast to the task of blur discrimination.

An alternative explanation is suggested here for the spatiotemporal receptive fields of Burr et al. In this explanation the receptive fields are not the receptive fields of actual single mechanisms, but instead descriptive functions that show how motion blur, i.e. the movement combined with the temporal filtering, modifies the appearance of the underlying spatial filter in space-time coordinates.

7.2. Linear shifter circuits

When proposing the shifter circuits principle for blur prevention of moving objects, Anderson and van Essen (1987) did not make any predictions of the amounts of blur their system would remove with different velocities. The results of the blur discrimination and the two-dot resolution experiments show that the velocity dependence of the effective blur produced by motion is linear. Can this kind of dependence be produced by shifter circuits? In principle, the shifter circuitry produces blur if the shift control is not accurate. In blur prevention, the shift control is guided by the velocity information. When the motion starts, it takes some time before the shift control returns a correct velocity value, and blur is produced. The amount of blur is the product of the delay time and the velocity, and if the delay is constant, the amount of blur depends linearly on the velocity. However, as soon as the velocity feedback signal reaches the right value, no more blur is produced. If the delay time is less than 50 ms and the motion summation time 100 ms, then the final internal image for our 150 ms presentation should be blur-free. If the sum of the delay time and the summation

time is more than 150 ms, some blur is left in the final image, and the amount of this blur relates linearly to the velocity. This, however, is an oversimplification of the situation. It does not take into account the fact that the feedback velocity signal has its own uncertainty, which also produces blur. If the error of the velocity signal is a linear function of velocity, then the blur produced by this uncertainty depends linearly on the velocity, as does also the total amount of blur. In reality, this is not necessarily the case. De Bruyn and Orban (1988) measured velocity discrimination thresholds for random dot patterns and moving light bars. At velocities of 1, 2, 4 and 8 deg/s, the Weber fractions for velocity discrimination were about 0.12, 0.09, 0.065 and 0.05, respectively - thus nowhere near a linear relation. If these thresholds reflect the behaviour of the hypothetical velocity feedback signal, as one would expect, then the amount of blur produced by the shifter circuits should not be a linear function of velocity. Since the amount of blur depends linearly on velocity, it is concluded that linear shifter circuits are not the mechanism behind the blur prevention.

7.3. Motion blur versus static spatial blur

Morgan and Benton (1989) assumed that motion introduces spatial blur. Their explanation for blur-free percepts of moving objects was that motion blur remains undetected because the internal representation of the object is already even more degraded by static spatial blur. The results of the blur discrimination experiments show that this is true up to velocities of about 2 deg/s. Morgan and Benton argued against any motion-deblurring mechanism, but the blur discrimination experiment of this thesis showed that the effective spatial blur produced by motion in a blur discrimination task is much less than predicted by a summation time of about 100 ms, indicating the existence of at least some kind of blur-free integration mechanism.

7.4. A new model for temporal integration

The essence of both the shifter circuit and the spatiotemporal receptive field principle is that motion blur is not produced because the mechanisms take into account the temporal delay at which different photoreceptors have been stimulated, i.e. integration follows the path of the moving object. Neither of these models is consistent with the results of experiments on random positional jitter. Badcock and Wong (1990a; 1990b) measured separation discrimination thresholds for two lines with a separation of 6 arc min, and found that positional jitter of up to 8 arc min had no effect on performance. Similar measurements were performed in this thesis, and the finding was confirmed but a temporal limitation was found: performance was resistant to positional noise only if the time interval between two consecutive stationary frames of the jittering target was at least 25 ms.

A new model for temporal integration is proposed that is also consistent with the findings of positional jitter experiments. In this model, the integration happens in two phases. The first phase summates information like a camera, forming a weighted average in time with weights defined by the temporal impulse response with a standard deviation of about 5 ms. Motion blur is produced in this phase only. The second phase integrates the images produced by the first phase over a time period of more than 100 ms in daylight. The second phase in translation-invariant: it uses a mechanism that can take into account changes of object location even if they are random. This phase does not produce motion blur. However, whatever the method of superimposing the information from the first phase images, spatial and luminance changes between the parts of the object lead to image degradation that may be difficult to discriminate from motion blur. The second phase does not remove blur either - in fact that would be dangerous since at that point the system could have no way of knowing what is blur in the image. Instead, the aim of this phase is to increase the signal-to-noise ratio or the statistical significance in discriminating the features in the image. In perceptual terms, this phase increases the perceived clarity of the image. The second phase integration in the model parallels very closely to the ideas of Dodwell (1971) who presented a model which is based on correlation of successive time

samples of retinal inputs, perceptual clarity being attained by a form of autocorrelation.

The model is also consistent with the results of Welch and McKee (1985) who found that if the two components of a vernier target are moving at the same velocity on converging paths, vernier thresholds are three to four times higher than those for a target with components travelling on parallel paths. According to the model, the first phase integration blurs the components similarly on both parallel and converging paths. However, the second phase degrades the target only if the spatial relations of the components change with time, i.e. when the components do not move as a rigid object in a rectilinear motion. Thus, performance can be expected to be poorer when the vernier components have converging trajectories.

Two basically different principles are suggested here to carry out the second phase integration, depending on the form of the final internal representation. In the first one, the final integrated image would be presented in topographical or retinotopic form, preserving neighbourhood relations. This would require that the shift in position of a moving object between a so far integrated image and a new first-phase output image should be first calculated, i.e. a correspondence problem should be solved. To do this accurately, the cross-correlation of the images would probably be needed. After shifting the object, the images could be added together. The shifting operation would require similar dynamic control as proposed in the linear shifter circuit principle of Anderson and van Essen (1987), but due to the need for a cross-correlation or other equally efficient method for the calculation of the shift, the total operation would be more complicated. If there were multiple separately moving objects in the image, or if the background was moving, the system would have to make an individual shift for each of them. It is argued here that this principle contains a causal inconsistency: to produce accurate shifts, the system should be able to analyse the image in detail, and if the system can do this, there is no need for shifts. The same argument applies also to the original linear shifter principle. Therefore, it is not very likely that this principle would underlie the second phase integration.

In the second solution, the integration is carried out by cross-correlation, i.e. by the basic operation of displacement mapping, and the final image is presented in a correlation form. A correlation image is a non-topographical representation that

presents all the positional relations inside the image, but not the absolute locations of the objects or their parts. One might argue that correlation does produce blur since variances are additive in correlation in the same way as they are in convolution. This effect, however, is cancelled out by the fact that the distance cues are also added together - in this respect correlation is just a scaling operation. Thus, correlation does not produce blur that would change the results of statistical detection or discrimination processes. Our models for blur discrimination presented in Eqs. (4.7), (4.8) and (4.9) are also consistent with the second phase integration based on correlation. Let us examine the model derivation in this case. If we assume that the internal representation of the edge, i.e. B' in Eq. (4.1), is not just the convolution of the real edge and the filters but the autocorrelation of that convolution, then the right side of the equation must be multiplied by $\sqrt{2}$. The same change must be made in Eqs. (4.3) and (4.4). Eq. (4.2) remains unchanged. The factor $\sqrt{2}$ cancels out in Eq. (4.5), and thus the rest of the model derivation and the final models presented in Eqs. (4.7), (4.8) and (4.9) remain the same even if the correlation phase is included in the model derivation.

Autocorrelation and cross-correlation have long played an important role in the computer applications of image processing. Anderson (1968) extended these principles to biological systems when he described a generalized memory storage model that utilizes spatial correlation functions. He showed mathematically the advantages of these principles in recognition, reconstruction and association of patterns. If our visual system was using this kind of memory, some of the tools that would be needed for the memory operations could also be used in the correlation-like temporal integration. In fact, the main functions of the correlation-like temporal integration might be memory-related. For instance, the spatial cross-correlation of two images of a stationary object, which is equivalent to the spatial autocorrelation, transforms the image to a translation-invariant form of representation that could be stored in the memory, or used as an address to retrieve a pattern from the memory if the memory is associative. On the other hand, if on the basis of some a priori information, a certain pattern from the memory were chosen as a starting image, the correlation-like integration would provide a measure of the degree of similarity between a new image and the stored pattern. If the new pattern were similar to the stored one, the detectability of the new pattern

would increase. The cross-correlation would also provide information about the differences between the old and new patterns.

If the visual system were using strategies such as auto-correlation and cross-correlation for signal energy detection, detectability would be directly related to signal-to-noise ratio. There has been a number of experiments demonstrating this type of relation (Burgess, 1990). Uttal (1975) made detection experiments with patterns of dots embedded in pictures of noise dots, and developed an autocorrelation theory for form detection to explain his results. There has also been one study supporting specifically the cross-correlation capability of humans. Burgess and Ghandeharian (1984) designed a model-free test for detection using a two-cycle sine-wave pattern. In some of their trials subjects were given complete knowledge and in some trials no knowledge of signal phase. The results of the phase-known trials were superior to the very best possible results for the phase-unknown trials. Burgess and Ghandeharian concluded that well-trained humans can perform cross-correlation detection when sufficient a priori information is available. One can interpret these results to show that cross-correlation is a strong candidate for a mechanism of temporal integration, and that the ability to use prior information or memory is an integral part of the mechanism.

The suggested model for temporal integration is not a motion-deblurring model in the original meaning of the word since it does not take into account the temporal delay at which different photoreceptors have been stimulated, nor does the integration follow the path of the moving object. To put it simply, the model is not motion-tuned. It explains the resistance of the visual system to positional noise when the random shifts of the target are introduced with time intervals longer than about 25 ms - and within a limited spatial window though the experiments does not allow one to make any predictions of the size of this window. The essence of the model and the interpretation of the findings in this study is that the form of a moving object is determined by using the same or similarly sized spatial filter or filters as for stationary objects up to velocities of at least 8 deg/s. Neurobiologically, the high spatial resolution found in the blur discrimination experiment is consistent with the idea implicit for instance in the paper of Livingstone and Hubel (1987) that the parvocellular-interblob stream is involved in high-resolution form perception, even with moving targets.

8. Summary and conclusions

Special motion-deblurring mechanisms have been suggested as an explanation for the absence of motion degradation in the percepts of moving objects. In this work, properties of visual motion blur were investigated in relation to these mechanisms and other possible explanations.

To separate the effects of motion blur and static spatial blur, blur discrimination thresholds for moving, Gaussian-blurred edges were measured. The results showed that motion does produce equivalent spatial blur. In theory, the source of this blur could be either camera-like summation, or the use of larger spatial filters for moving than for stationary objects, or a mixture of these two. To estimate the amount of the blur, a mathematical model based on a linear transform of the physical image of the edge to its neural representation was derived. The velocity dependence of the equivalent blur was found to be linear. This implied that camera-like blur alone is sufficient to explain the results, but it did not rule out other types of blur with a linear velocity dependence. The extent of the equivalent blur could be predicted by a temporal impulse response with a standard deviation of about 5 ms, suggesting that the blur-producing motion integration time is about 20 to 25 ms. When the direction of motion was made perpendicular to the edge profile or to the separation of dots in a two-dot resolution experiment, velocity had no detrimental effect on the thresholds, supporting strongly the view that motion blur results from camera-like summation.

When one observer repeated the blur discrimination experiment after six months, changes in thresholds and model parameters were found that indicate low-level learning and neural plasticity in the blur discrimination system. However, when the contrast polarity in relation to the direction of motion was changed, the learning effects were not transferred. This indicates that spatial analysis of objects in motion could be served by two separate subsystems which may be related to the on and off systems of the visual pathway.

The amount of motion blur estimated in the blur discrimination experiment corresponds to only about a fifth of the total 100 ms temporal integration time in the detection of moving objects. If

the visual system uses the same mechanism for the detection and the spatial analysis of a moving object, then an integration stage that does not produce blur is needed to explain the difference. The findings of the blur discrimination experiments did not provide any direct evidence either for or against the special motion-deblurring models: the shifter circuit principle and the spatiotemporal receptive field model. However, indirect evidence against and alternative explanations were found. Most importantly, both of the models are motion-tuned, and neither is resistant to random positional jitter, a property of the visual system that was confirmed in this work in a spatial interval discrimination experiment.

A new model for temporal integration is proposed to explain the resistance of the visual system to positional noise. This model integrates spatial information in two phases. The first phase is a camera-like exposure phase governed by the temporal impulse response of the system. This phase produces motion blur. The second phase integrates the images produced by the first phase using a translation-invariant mechanism. This phase does not produce or remove motion blur. Instead, it increases the signal-to-noise ratio of the image. The most important point where this model differs from the special motion-deblurring models is that it is not motion-tuned. It is proposed that the second phase mechanism could be based on cross-correlation of successive images. A related model that uses correlation principles to form a simultaneous representation of the shape and the velocity of a moving object is separately proposed under the name of displacement mapping.

The results of this study indicate that temporal integration does not have to be motion-tuned to explain human performance in spatial discriminations of moving objects. In fact, it is shown that a non-motion-tuned model can explain a broader range of phenomena than a motion-tuned model. A question not yet convincingly answered is whether the total integration time of about 100 ms is used at all in the spatial shape analysis of moving objects, or whether it just reflects the summation of information for the detection and analysis of motion.

References

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A 2, 284-299.

Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature* **300**, 523-525.

Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science* **255**, 1141-1143.

Anderson, C. H. & van Essen, D. C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proc. Natl. Acad. Sci. USA* **84**, 6297-6301.

Anderson, J. A. (1968). A memory storage model utilizing spatial correlation functions. *Kybernetik* **5**, 113-119.

Anderson, S. J. & Burr, D. C. (1985). Spatial and temporal frequency selectivity of the human motion detection system. *Vision Res.* **25**, 1147-1154.

Anderson, S. J. & Burr, D. C. (1987). Receptive field size of human motion detector units. *Vision Res.* **27**, 621-635.

Anderson, S. J. & Burr, D. C. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Res.* **29**, 1343-1358.

Anderson, S. J. & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *J. Opt. Soc. Am. A* **8**, 1330-1339.

Anderson, S. J., Burr, D. C. & Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *J. Opt. Soc. Am. A* **8**, 1340-1351.

Anstis, S. (1983). Visual coding of position and motion. In *Physical and Biological Processing of Images* (ed. Braddick, O. J. & Sleigh, A. C.). Berlin/Heidelberg/New York: Springer Verlag.

Badcock, D. R. & Wong, T. L. (1990a). Resistance to positional noise in human vision. *Nature* **343**, 554-555.

Badcock, D. R. & Wong, T. L. (1990b). The sensitivity of separation discrimination to spatiotemporal jitter. *Vision Res.* **30**, 1555-1560.

Baker, C. L. Jr & Braddick, O. (1985). Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Res.* **25**, 803-812.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *J. Physiol.* **141**, 337-350.

Barlow, H. B. (1981). Critical limiting factors in the design of the eye and visual cortes. *Proc. R. Soc. Lond. B* **212**, 1-34.

Barlow, H. B. & Levick, R. W. (1965). The mechanism of directionally selective units in rabbit's retina. *J. Physiol.* **173**, 377-407.

Barlow, H. B. & Levick, W. R. (1969a). Changes in the maintained discharge with adaptation level in the cat retina. *J. Physiol.* **202**, 699-718.

Barlow, H. B. & Levick, W. R. (1969b). Three factors limiting the reliable detection of light by retinal ganglion cells in the cat. *J. Physiol.* **200**, 1-24.

Baylor, D. A., Hodgkin, A. L. & Lamb, T. D. (1974). Reconstruction of the electrical responses of turtle cones to flashes and steps of light. *J. Physiol.* **242**, 759-791.

Bergen, J. R. & Wilson, H. R. (1985). Prediction of flicker sensitivities from temporal three-pulse data. *Vision Res.* **25**, 577-582.

Bergen, J. R., Wilson, H. R. & Cowan, J. D. (1979). Further evidence for four mechanisms mediating vision at threshold: sensitivities to complex gratings and aperiodic stimuli. *J. Opt. Soc. Am.* **69**, 1580-1587.

Bischof, W. F. & Di Lollo, V. (1990). Perception of directional sampled motion in relation to displacement and spatial frequency: evidence for a unitary motion system. *Vision Res.* **30**, 1341-1362.

Blakemore, C. & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.* **203**, 237-260.

Bloch (1885). Experience sur la vision. *C. r. Séanc. Soc. Biol.* **37**, 493-495.

Borst, A. & Egelhaaf, M. (1989). Principles of visual motion detection. *Trends Neurosci.* **12**, 297-306.

Boulton, J. C. & Baker, C. L. Jr (1991). Motion detecion is dependent on spatial frequency not size. *Vision Res.* **31**, 77-87.

Bowne, S. F., McKee, S. & Glaser, D. A. (1989). Motion interference in speed discrimination. *J. Opt. Soc. Am. A* **6**, 1112-1121.

Braddick, O. (1974). Short range process in apparent motion. *Vision Res.* **25**, 839-847.

Bülthoff, H., Little, J. & Poggio, T. (1989). A parallel algorithm for real-time computation of optical flow. *Nature* **337**, 549-553.

Burbeck, C. A. & Kelly, D. H. (1980). Spatiotemporal characteristics of visual mechanisms: excitory-inhibitory model. *J. Opt. Soc. Am.* **70**, 1121-1126.

Burbeck, C. A. & Kelly, D. H. (1981). Contrast gain measurements and the transient/sustained dichotomy. *J. Opt. Soc. Am.* **71**, 1335-1342.

Burgess, A. E. (1990). High level visual decision efficiences. In *Vision: Coding and Efficiency* (ed. Blakemore, C.). Cambridge: Cambridge University Press.

Burgess, A. E. & Ghandeharian, H. (1984). Visual signal detection I: phase sensitive detection. *J. Opt. Soc. Am. A* **1**, 900-905.

Burr, D. C. (1979). Acuity for apparent vernier offset. *Vision Res.* **19**, 835-837.

Burr, D. C. (1980). Motion smear. *Nature* **284**, 164-165.

Burr, D. C. (1981). Temporal summation of moving images by the human visual system. *Proc. R. Soc. Lond. B* **211**, 321-339.

Burr, D. C. (1991). Human sensitivity to flicker and motion. In *Limits of Vision* (ed. Kulikowski, J. J., Walsh, V. & Murray, I. J.). London: Macmillan.

Burr, D. C. & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Res.* **22**, 479-484.

Burr, D. C. & Ross, J. (1986). Visual processing of motion. *Trends Neurosci.* **9**, 304-307.

Burr, D. C., Ross, J. & Morrone, M. C. (1986a). Seeing objects in motion. *Proc. R. Soc. Lond. B* **227**, 249-265.

Burr, D. C., Ross, J. & Morrone, M. C. (1986b). Smooth and sampled motion. *Vision Res.* **26**, 645-652.

Burt, P. & Sperling, G. (1981). Time distance and feature trade-offs in visual apparent motion. *Psychol. Rev.* **88**, 171-195.

Campbell, F. W., Cooper, G. F. & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *J. Physiol.* **203**, 223-235.

Campbell, F. W. & Gubisch, R. W. (1966). Optical quality of the human eye. *J. Physiol.* **186**, 558-578.

Campbell, F. W., Nachmias, J. & Jukes, J. (1970). Spatial frequency discrimination in human vision. *J. Opt. Soc. Am.* **60**, 555-559.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *J. Physiol.* **197**, 551-566.

Cavanagh, P., Boeglin, J. & Favreau, O. E. (1985). Perception of motion in equiluminous kinematograms. *Perception* **14**, 151-162.

Cavanagh, P. & Mather, G. (1989). Motion: The long and short of it. *Spatial Vision* **4**, 103-129.

Chang, J. J. & Julesz, B. (1983). Displacement limits for spatial frequency filtered rando-dot cinematograms in apparent motion. *Vision Res.* **23**, 1379-1385.

Chino, Y. M., Kaas, J. H., Smith III, E. L., Langston, A. L. & Cheng, H. (1992). Rapid reorganization of cortical maps in adult cats following restricted deafferentation in retina. *Vision Res.* **32**, 789-796.

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A* **5**, 1986-2007.

Cleary, R. (1990). Contrast dependence of short-range apparent motion. *Vision Res.* **30**, 463-478.

Cleland, B. G., Harding, D. H. & Tulunay-Keesey, U. (1979). Visual resolution and receptive field size: Examination of two kinds of cat retinal cell. *Science* **205**, 1015-1017.

Cohen, M. A. & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion and resonance. *Percept. Psychophys.* **36**, 428-456.

Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.* **27**, 183-228.

Cornsweet, T. N. (1970). *Visual Perception*. New York: Academic Press.

Dawson, M. & Di Lollo, V. (1990). Effects of adapting luminance and stimulus contrast on the temporal spatial limits of short-range motion. *Vision Res.* **30**, 415-429.

De Bruyn, B. & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Res.* **28**, 1323-1335.

De Lange, H. (1952). Experiments on flicker and some calculations of an electrical analogue of the foveal system. *Physica* **18**, 935-950.

De Lange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and coloured light. *J. Opt. Soc. Am.* **48**, 777-784.

De Valois, R. L., Albrecht, D. G. & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545-559.

Derrington, A. & Suero, M. (1991). Motion of complex patterns is computed from the perceived motions of their components. *Vision Res.* **31**, 139-149.

Derrington, A. M. & Lennie, P. (1982). The influence of temporal frequency and adaptation level on receptive field organisation of retinal ganglion cells in cat. *J. Physiol.* **333**, 343-366.

Derrington, A. M. & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurons in the lateral geniculate nucleus of macaque. *J. Physiol.* **357**, 219-240.

DeYoe, E. A. & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* **11**, 219-226.

Dodwell, P. C. (1971). On perceptual clarity. *Psychol. Rev.* **78**, 275-289.

Douglas, R. J. & Martin, K. A. C. (1991). Opening the grey box. *Trends Neurosci.* **14**, 286-293.

Dreher, B. & Sanderson, K. J. (1973). Receptive field analysis: responses to moving visual contours by single lateral geniculate neurons in the cat. *J. Physiol.* **234**, 95-118.

Enroth-Cugell, C. & Robson, J. G. (1966). The contrast sensitivity function of retinal ganglion cells in the cat. *J. Physiol.* **187**, 517-552.

Enroth-Cugell, C., Robson, J. G., Schweitzer-Tong, D. E. & Watson, A. B. (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *J. Physiol.* **341**, 279-307.

Fahle, M. & Poggio, T. (1981). Visual hyperacuity: Spatiotemporal interpolation in human vision. *Proc. R. Soc. Lond. B* **213**, 451-477.

Ferrera, V. P. & Wilson, H. R. (1990). Perceived direction of two-dimensional moving pattern. *Vision Res.* **30**, 273-287.

Ferrera, V. P. & Wilson, H. R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Res.* **31**, 877-893.

Ferry, E. S. (1892). Persistence of vision. *Am. J. Sci.* **44**, 192.

Finney, D. J. (1971). *Probit Analysis*. Cambridge: Cambridge Univ. Press.

Fleet, D., Hallett, P. E. & Jepson, A. D. (1985). Spatiotemporal inseparability in early visual processing. *Biol. Cybern.* **52**, 153-164.

Gibson, J. J. & Gibson, E. J. (1957). Continuous perceptive transformations and the perception of rigid motion. *J. Exp. Psychol.* **54**, 129-138.

Gilbert, C. D. & Wiesel, T. N. (1991). Reorganization of topography of macaque primary visual cortex following focal retinal lesions. *Invest. Ophtalmol. Vis. Sci.* **32**, 1952.

Gizzi, M. S., Katz, E., Schumer, R. A. & Movshon, J. A. (1990). Selectivity for orientation and direction of motion of single neurons in cat striate and extrastriate visual cortex. *J. Neurophysiol.* **63**, 1529-1543.

Gottsdanker, R. M. (1956). The ability of human operators to detect acceleration of target motion. *Psychol. Bull.* **53**, 477-487.

Gouras, P. & Zrenner, E. (1981). Color coding in primate retina. *Vision Res.* **21**, 1591-1598.

Graham, C. H. & Margaria, R. (1935). Area and intensity-time relation in the peripheral retina. *Am. J. Physiol.* **113**, 299-305.

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. *Vision Res.* **11**, 251-259.

Greenlee, M. W., Gerling, J. & Waltenspiel, S. (1990). Spatial-frequency discrimination of drifting gratings. *Vision Res.* **30**, 1331-1339.

Gregory, R. L. (1990). What is caught in neural nets (editorial). *Perception* **19**, 561-568.

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color and brightness perception. I: Monocular theory. *Percept. Psychophys.* **41**, 87-116.

Grossberg, S. & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* **92**, 173-211.

Grossberg, S. & Rudd, M. E. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychol. Rev.* **99**, 78-121.

Grossberg, S. & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Percept. Psychophys.* **43**, 241-277.

Hamerly, J. R. & Dvorak, C. A. (1981). Detection and discrimination of blur in edges and lines. *J. Opt. Soc. Am.* **71**, 448-452.

Harris, M. G. (1986). The perception of moving stimuli: a model of spatiotemporal coding in human vision. *Vision Res.* **26**, 1281-1287.

Hayhoe, M. M. (1990). Spatial interactions and models of adaptation. *Vision Res.* **30**, 957-965.

Hayhoe, M. M., Benimoff, N. I. & Hood, D. C. (1987). The time-course of multiplicative and subtractive adaptation process. *Vision Res.* **27**, 1981-1996.

Hess, R. F. & Plant, G. T. (1985). Temporal frequency discrimination in human vision: evidence for an additional mechanism in the low spatial and high temporal frequency region. *Vision Res.* **25**, 1493-1500.

Hildreth, E. C. (1984). *The Measurement of Visual Motion*. Cambridge: MIT Press.

Hildreth, E. C. & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Ann. Rev. Neurosci.* **10**, 477-533.

Hochstein, S. & Shapley, R. M. (1976). Linear and non-linear spatial subunits in Y cat retinal ganglion cells. *J. Physiol.* **262**, 265-285.

Horn, B. K. P. & Schunk, B. G. (1981). Determining optical flow. *Artif. Intell.* **17**, 185-203.

Howell, E. R. & Hess, R. F. (1978). The functional area for summation to threshold for sinusoidal gratings. *Vision Res.* **18**, 369-374.

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *J. Physiol.* **148**, 754-791.

Hubel, D. H. & Wiesel, T. N. (1960). Receptive fields of optic nerve fibres in the spider monkey. *J. Physiol.* **154**, 572-580.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106-154.

Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**, 215-243.

Ikeda, H. & Wright, M. J. (1975). Spatial and temporal properties of "sustained" and "transient" neurones in area 17 of the cat's visual cortex. *Exp. Brain Res.* **22**, 362-382.

Ikeda, M. (1965). Temporal summation of positive and negative flashes in the visual system. *J. Opt. Soc. Am.* **55**, 1527-1534.

Ikeda, M. (1986). Temporal impulse response. *Vision Res.* **26**, 1431-1440.

Johannson, G. (1975). Visual motion perception. *Sci. Am.* **232**, 76-88.

Kaas, J. H., Krubitzer, L. A., Chino, Y. M., Langston, A. L., Polley, E. H. & Blair, N. (1990). Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science* **248**, 229-231.

Kaplan, E. & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Pro. Natl. Acad. Sci. USA* **83**, 2755-2757.

Keesey, U. T. (1971). Flicker and pattern detection: A comparison of thresholds. *J. Opt. Soc. Am.* **62**, 60-62.

Kelly, D. H. (1961). Visual response to time dependent stimuli. I. Amplitude sensitivity measurements. *J. Opt. Soc. Am.* **51**, 422-429.

Kelly, D. H. (1971). Theory of flicker and transient responses, I. Uniform fields. *J. Opt. Soc. Am.* **61**, 537-546.

Kelly, D. H. (1972). *Flicker*. Berlin: Springer-Verlag.

Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.* **69**, 1340-1349.

Kelly, D. H. & Burbeck, C. A. (1987). Further evidence for a broadband, isotropic mechanism sensitive to high-velocity stimuli. *Vision Res.* **27**, 1527-1537.

Kelly, D. H. & Savoie, R. E. (1978). Theory of flicker and transient responses. III. An essential nonlinearity. *J. Opt. Soc. Am.* **68**, 1481-1490.

King-Smith, P. E. & Kulikowski, J. J. (1975). The detection of gratings by independent activation of line detectors. *J. Physiol.* **247**, 237-271.

Kingdom, F. & Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Res.* **32**, 1565-1582.

Klein, S. A. & Levi, D. M. (1985). Hyperacuity thresholds of 1.0 second: Theoretical predictions and empirical validation. *J. Opt. Soc. Am. A* **2**, 1170-1190.

Kolers, P. A. (1972). *Aspects of Motion Perception*. New York: Pergamon Press.

Kooi, F. L., Grosof, D.H., DeValois, K. K. & DeValois, R. L. (1988). *Optical Society of America, Technical Digest Series* **17**, THS3.

Kuffler, S. W. (1953). Discharge patterns and functional reorganization of mammalian retina. *J. Neurophysiol.* **16**, 37-68.

Kulikowski, J. J. (1971). Effects of eye movements on the contrast sensitivity of spatio-temporal patterns. *Vision Res.* **11**, 261-273.

Kulikowski, J. J. (1991). What really limits vision? Conceptual limitations to the assessment of visual function and the role of interacting channels. In *Limits of Vision* (ed. Kulikowski, J. J., Walsh, V. & Murray, I. J.). London: Macmillan.

Kulikowski, J. J. & Bishop, P. O. (1981). Linear analysis of the responses of simple cells in the cat visual cortex. *Exp. Brain Res.* **44**, 386-400.

Kulikowski, J. J. & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient neurons in the human visual system. *J. Physiol.* **232**, 149-162.

Lappin, J. S. & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Res.* **16**, 161-168.

Lee, B. B., Elephandt, A. & Virsu, V. (1981). Phase of responses to sinusoidal gratings of simple cells in cat striate cortex. *J. Neurophysiol.* **45**, 818-828.

Legge, G. E. & Foley, J. M. (1980). Contrast masking in human vision. *J. Opt. Soc. Am.* **70**, 1458-1470.

Levi, D. M. & Klein, S. A. (1986). Sampling in spatial vision. *Nature* **320**, 360-362.

Levi, D. M. & Klein, S. A. (1990). Equivalent intrinsic blur in spatial vision. *Vision Res.* **30**, 1971-1993.

Levinson, E. & Sekuler, R. (1975). The independence of channels in human vision selective for direction of movement. *J. Physiol.* **250**, 347-366.

Levinson, J. Z. (1968). Flicker fusion phenomena. *Science* **160**, 21-28.

Livingstone, M. S. & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement and depth. *J. Neurosci.* **7**, 3416-3468.

Maffei, L. & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyzer. *Vision Res* **13**, 1255-1267.

Mandler, M. B. (1984). Temporal frequency discrimination above threshold. *Vision Res.* **24**, 1873-1880.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D. & Hildreth, E. C. (1980). Theory of edge detection. *Proc. R. Soc. Lond. B* **207**, 187-217.

Marr, D. & Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Research Program Bulletin* **15**, 470-488.

Marr, D., Poggio, T. & Hildreth, E. (1980). Smallest channel in early human vision. *J. Opt. Soc. Am.* **70**, 868-871.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B* **211**, 151-180.

Mather, G. (1987). The dependence of edge displacements thresholds on edge blur, contrast, and displacement distance. *Vision Res.* **27**, 1631-1637.

Mather, G. (1988). Temporal properties of apparent motion in subjective figures. *Perception* **17**, 729-736.

Maunsell, J. H. R. & van Essen, D. C. (1983). Functional properties of middle temporal area of the macaque monkey. I. Selectivity for stimulus direction, speed and orientation. *J. Neurophysiol.* **49**, 1127-1147.

McKee, S. P. & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Res.* **1984**, 25-32.

McKee, S. P., Silverman, G. H. & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Res.* **26**, 609-619.

McKee, S. P. & Taylor, D. G. (1984). Discrimination of time: comparison of foveal and peripheral sensitivity. *J. Opt. Soc. Am. A* **1**, 620-627.

Meijer, J. G., van der Wildt, G. & van den Brink, G. (1978). Twin-flash response as a function of flash diameter. *Vision Res.* **18**, 1111-1116.

Morgan, M. J. (1976). Pulfrich effect and the filling in of apparent motion. *Perception* **5**, 187-195.

Morgan, M. J. (1979). Perception of continuity in strobocopic motion: a temporal frequency analysis. *Vision Res.* **19**, 491-500.

Morgan, M. J. (1980a). Analog models of motion perception. *Proc. R. Soc. Lond. B* **290**, 117-135.

Morgan, M. J. (1980b). Spatiotemporal filtering and the interpolation effect in apparent motion. *Perception* **9**, 161-174.

Morgan, M. J. (1985). *Computational vision.* London: British Psychology Society.

Morgan, M. J. (1991). Hyperacuity. In *Spatial Vision* (ed. Regan, D.). London: Macmillan.

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature* **355**, 344-346.

Morgan, M. J. & Benton, S. (1989). Motion-deblurring in human vision. *Nature* **340**, 385-386.

Morgan, M. J., Ross, J. & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biol. Cybern.* **65**, 113-119.

Morgan, M. J. & Ward, R. M. (1985). Spatial and spatial-frequency primitives in spatial-interval discrimination. *J. Opt. Soc. Am. A* **2**, 1205-1210.

Morgan, M. J. & Watt, R. J. (1982). The modulation transfer function of a display oscilloscope: measurements and comments. *Vision Res.* **22**, 1083-1085.

Morgan, M. J. & Watt, R. J. (1983). On the failure of spatiotemporal interpolation: a filtering model. *Vision Res.* **23**, 997-1004.

Morgan, M. J., Watt, R. J. & McKee, S. P. (1983). Exposure duration affects the sensitivity of vernier acuity to target motion. *Vision Res.* **23**, 541-546.

Moulden, B. & Begg, H. (1986). Some tests of the Marr-Ullman model of movement detection. *Perception* **15**, 139-155.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1985). The analysis of moving visual patterns. In *Pattern Recognition Mechanisms* (ed. Chagas, C., Gattas, R. & Gross, C. G.). Rome: Vatican Press.

Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 an 18 of the cat's visual cortex. *J. Physiol.* **283**, 101-120.

Nachmias, J. & Sansbury, R. V. (1974). Grating contrast discrimination may be better than detection. *Vision Res.* **14**, 1039-1042.

Nakayama, K. (1985). Biological image motion processing: a review. *Vision Res.* **25**, 625-660.

Nakayama, K. & Silverman, G. H. (1985). Detection and discrimination of sinusoidal grating displacements. *J. Opt. Soc. Am. A* **2**, 267-274.

Nakayama, K. & Tyler, C. W. (1981). Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Res.* **21**, 427-433.

Nielsen, K. R. K., Watson, A. B. & Ahumada, A. J. (1985). Application of a computable model of human spatial vision to phase discrimination. *J. Opt. Soc. Am. A* **2**, 1600-1606.

Oppenheim, A. V. & Lim, J. S. (1981). The importance of phase in signals. *Proc IEEE* **69**, 529-541.

Piotrowski, L. N. & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception* **11**, 337-346.

Poggio, T. (1983). *Visual algorithms*. New York: Springer-Verlag.

Poggio, T., Fahle, M. & S., E. (1992). Fast perceptual learning in visual hyperacuity. *Science* **256**, 1018-1021.

Poggio, T. & Reichardt, W. (1973). Considerations on models of movement detection. *Kybernetik* **13**, 223-227.

Porter, T. C. (1902). Contributions to the study of flicker. *Proc. Rev. Soc.* **17**, 313-329.

Press, I. & William, H. (1988). *Numerical Recipes in C*. Cambridge: Cambridge University Press.

Purkinje, J. (1825). *Beobachtungen and Versuche zur Physiologie der Sinne*. Bd II, 60.

Ramachandran, V. S. (1990). Interactions between motion, depth, color and form: the utilitarian theory of perception. In *Vision: Coding and Efficiency* (ed. Blakemore, C.). Cambridge: Cambridge University Press.

Ramachandran, V. S. & Anstis, S. M. (1985). Perceptual organization in multistable apparent motion. *Perception* 135-143.

Ramachandran, V. S. & Cavangh, P. (1987). Motion capture anisotrophy. *Vision Res.* **27**, 97-106.

Ramachandran, V. S. & Inada, V. (1985). Spatial phase and frequency in motion capture of random-dot patterns. *Spatial Vision* **1**, 57-67.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Res.* **13**, 1399-1401.

Regan, D. (1989). *Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine*. New York: Elsevier.

Regan, D. (1991). Stimuli and methods in spatial vision research. In *Spatial Vision* (ed. Regan, D.). London: Macmillan.

Reichardt, W. (1957). Autokorrelationsauswetung als Funktionsprinzip des Zentralnervensystems. *Z. Naturforch. Teil B* **12**, 447-457.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory Communication* (ed. Rosenblith, W. A.). New York: Wiley.

Reichardt, W. (1969). Movement perception in insects. In *Processing of Optical Data by Organisms and Machines* (ed. Reichardt, W.). London/New York: Academic Press.

Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *J. Opt. Soc. Am.* **56**, 1141-1142.

Robson, J. G. (1975). Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. In *Handbook of Perception, Vol. 5* (ed. Carterette, E. C. & Friedman, M. P.). New York: Academic Press.

Robson, J. G. & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res.* **21**, 409-418.

Rodieck, R. W. & Stone, J. (1965). Response of cat retinal ganglion cells to moving visual patterns. *J. Neurophysiol.* **28**, 819-832.

Roufs, J. A. J. & Blommaert, F. J. J. (1981). Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique. *Vision Res.* **21**, 1203-1221.

Rovamo, J. & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Exp. Brain Res.* **37**, 495-510.

Schiller, P. H. & Logothetis, N. K. (1990). The color-opponent and broad-band channels in the primate visual system. *Trends Neurosci.* **10**, 392-398.

Schmerler, J. (1976). The visual perception of accelerated motion. *Perception* **5**, 187-195.

Shapley, R. & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Ann. Rev. Neurosci.* **8**, 547-583.

Shapley, R. & Perry, V. H. (1986). Cat and monkey retinal ganglion cells and visual functional roles. *Trends Neurosci.* 229-335.

Snowden, R. J. & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Res.* **31**, 907-914.

Stigmar, G. (1971). Blurred visual stimuli. II. The effect of blurred visual stimuli on Vernier and stereo acuity. *Acta Opthalmol.* **49**, 364-379.

Stoner, G. R., Albright, T. D. & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature* **344**, 153-155.

Stromeyer, C. F. & Klein, S. (1975). Evidence against narrow-band spatial frequency channels in human vision: the detectability of frequency modulated gratings. *Vision Res.* **15**, 899-910.

Thompson, P. (1983). Discrimination of moving gratings at and above detection threshold. *Vision Res.* **23**, 1533-1538.

Toet, A. & Koenderink, J. J. (1987). Two-point discrimination at low resolution. *J. Opt. Soc. Am. A* **4**, 1448-1454.

Toet, A., van Eekhout, M. P., Simons, H. L. J. J. & Koenderink, J. J. (1987). Scale invariant features of differential spatial displacement discrimination. *Vision Res.* **27**, 441-451.

Tolhurst, D. J. (1973). Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *J. Physiol.* **231**, 385-402.

Tolhurst, D. J. & Movshon, J. A. (1975). Spatial and temporal contrast sensitivity of striate cortical neurones. *Nature* **257**, 674-675.

Ueno, T. (1977). Temporal characteristics of the human visual system as revealed by reaction time to double pulses of light. *Vision Res.* **17**, 591-596.

Uetsuki, T. & Ikeda, M. (1970). Study of temporal visual response by summation index. *J. Opt. Soc. Am.* **60**, 377-381.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, Mass.: The MIT Press.

Ullman, S. (1983). The measurement of visual motion. Computational considerations and some neurophysiological implications. *Trends Neurosci.* **6**, 177 - 179.

Ullman, S. (1986). Artificial intelligence and the brain: Computational studies of the visual system. *Ann. Rev. Neurosci.* **9**, 1-26.

Ullman, S. & Hildreth, E. C. (1983). The measurement of visual motion. In *Physical and Biological Processing of Images* (ed. Braddick, O. J. & Sleigh, A. C.). Berlin/Heidelberg/New York: Springer Verlag.

Uttal, W. R. (1975). *An Autocorrelation Theory of Form Detection*. Hillsdale, New Jersey: Erlbaum.

van de Grind, W. A., Koenderink, J. J. & van Doorn, A. J. (1987). Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *J. Opt. Soc. Am. A* **4**, 1643-1652.

van Essen, D. C., Anderson, C. H. & Felleman, D. J. (1992). Information processing in the primate visual sustem: an integrated systems perspective. *Science* **255**, 419-423.

Van Ness, F. L., Koenderink, J. J., Nas, H. & Bouman, M. H. (1967). Spatio-temporal modulation transfer in the human eye. *J. Opt. Soc. Am.* **57**, 401-406.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *J. Opt. Soc. Am. A* **1**, 451-473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300-321.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *J. Exp. Psychol.* **45**, 205-217.

Watson, A. B. (1979). Probability summation over time. *Vision Res.* **19**, 515-522.

Watson, A. B. (1982). Derivation of the impulse response: Comments on the method of Roufs and Blommaert. *Vision Res.* **22**, 1335-1337.

Watson, A. B. & Ahumada, A. J. (1983). A look at motion in the frequency domain. *Nasa Tech. Memo.* **84352**

Watson, A. B. & Ahumada, A. J. (1985). Model of human visual motion sensing. *J. Opt. Soc. Am. A* **2**, 322-342.

Watson, A. B. & Nachmias, J. (1977). Patterns of temporal interaction in the detection of gratings. *Vision Res.* **17**, 893-90.

Watson, A. B. & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Res.* **21**, 1115-1122.

Watson, A. B., Thompson, P. G., Murphy, B. J. & Nachmias, J. (1980). Summation and discrimination of gratings moving in opposite directions. *Vision Res.* **20**, 341-347.

Watson, A. B., Ahumada, A. J. & Farrell, J. F. (1980). Window of visibility: A psychophysical theory of fidelity in time-sampled displays. *J. Opt. Soc. Am. A* **3**, 300-307.

Watt, R. J. (1984). Towards a general theory of the visual acuitites for shape and spatial arrangement. *Vision Res.* **24**, 1377-1386.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *J. Opt. Soc. Am. A* **4**, 2006-2021.

Watt, R. J. (1988). *Visual Processing*. Hove, East Sussex: Erlbaum.

Watt, R. J. & Andrews, D. P. (1981). APE: adaptive probit estimation of psychometric functions. *Curr. Psychol. Rev.* **1**, 205-214.

Watt, R. J. & Morgan, M. J. (1983a). Mechanisms responsible for the assessment of visual localization: Theory and evidence. *Vision Res.* **23**, 97-109.

Watt, R. J. & Morgan, M. J. (1983b). The recognition and representation of edge blur: evidence for spatial primitives in human vision. *Vision Res.* **23**, 1464-1477.

Watt, R. J. & Morgan, M. J. (1984). Spatial filters and the localization of luminance changes in human vision. *Vision Res.* **24**, 1387-1397.

Watt, R. J. & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Res.* **25**, 1661-1674.

Watt, R. J., Morgan, M. J. & Ward, R. M. (1983). The use of different cues in Vernier acuity. *Vision Res.* **23**, 991-995.

Welch, L. (1989). The perception of moving plaids reveal two motion-processing stages. *Nature* **337**, 734-736.

Welch, L. & Bowne, S. F. (1990). Coherence determines speed discrimination. *Perception* **19**, 425-435.

Welch, L. & McKee, S. P. (1985). Colliding targets: evidence for spatial localization within the motion system. *Vision Res.* **25**, 1901-1910.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.* **61**, 151-265.

Westheimer, G. (1975). Visual acuity and hyperacuity. *Invest. Ophtalmol. Vis. Sci.* **14**, 570-571.

Westheimer, G. (1979). The spatial sense of the eye. *Invest. Ophtalmol. Vis. Sci.* **17**, 893-912.

Westheimer, G. & McKee, S. P. (1975). Visual acuity in the presence of retinal image motion. *J. Opt. Soc. Am.* **65**, 847-850.

Westheimer, G. & McKee, S. P. (1977). Integration regions for visual hyperacuity. *Vision Res.* **17**, 89-93.

Westheimer, G. & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *J. Opt. Soc. Am.* **68**, 450-455.

Westheimer, G. & McKee, S. P. (1980). Stereoscopic acuity with defocused and spatially filtered images. *J. Opt. Soc. Am.* **70**, 772-778.

Whitten, D. N. & Brown, K. T. (1973). Slowed decay of the monkey's cone receptor potential by intense stimuli, and protection from this effect by light adaptation. *Vision Res.* **13**, 1659-1667.

Williams, R. A., Enoch, J. M. & Essock, E. A. (1984). The resistance of selected hyperacuity configurations to retinal image degradation. *Invest. Ophtalmol. Vis. Sci.* **25**, 389-399.

Wilson, H. R. (1978). Quantitative characterization of two-types of line-spread function near the fovea. *Vision Res.* **18**, 971-981.

Wilson, H. R. (1991). Psychophysical models of spatial vision and hyperacuity. In *Spatial Vision* (ed. Regan, D.). London: Macmillan.

Wilson, H. R. & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Res.* **19**, 19-32.

Wilson, H. R. & Gelb, D. J. (1984). Modified line element theory for spatial frequency and width dicrimination. *J. Opt. Soc. Am. A* **1**, 124-131.

Wilson, H. R., McFarlane, D. K. & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Res,* **23**, 871-882.

Yuille, A. L. & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature* **333**, 71-74.

Zemon, V., Gordon, J. & Welch, J. (1988). Asymmetries in On and Off visual pathways of human revealed using contrast-evoked cortical potentials. *Vis. Neurosci.* **1**, 145-150.